(1)

CONTRACT NO.: DAMD17-87-C-7188

TITLE: NEUROTOXIN AND EPITOPE STRUCTURAL STUDIES

PRINCIPAL INVESTIGATOR: DONALD F. HUNT

PI ADDRESS: University of Virginia
McCormick Road
Charlottesville, Virginia 22906

REPORT DATE: January 25, 1991

TYPE OF REPORT: Final

PREPARED FOR: U.S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND
FORT DETRICK
FREDERICK, MARYLAND 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: University of Virginia; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code): McCormick Road, Charlottesville, Virginia 22906	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Medical Research & Development Command; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: DAMD17-87-C-7188
8c. ADDRESS (City, State, and ZIP Code): Fort Detrick, Frederick, Maryland 21702-5012	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. 61102A; PROJECT NO. 3M1-61102BS12; TASK NO. AA; WORK UNIT ACCESSION NO. DA313304

11. TITLE (Include Security Classification)

(U) Neurotoxin and Epitope Structural Studies

12. PERSONAL AUTHOR(S)
Donald F. Hunt

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 9/28/87 TO 9/27/90	1991 January 25	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	RA 1; Botulinum neurotoxin; BD; Antigenic determinant; Tandem mass spectrometry
06	01		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Neurotoxins of botulinum clostridium are scientifically interesting for two reasons. First, they are extremely toxic. Second, they can be used as models for three important biological phenomena, selective recognition by a target cell, transport through the plasma membrane and toxic activity. All three activities are situated on one polypeptide of approx. 150 kDa. Whereas the complete gene sequences of neurotoxins A and C1 were published very recently only parts of the neurotoxin E sequence is known. By using mass spectrometry, supported by automated Edman degradation we were able to deduce approx. 50 kDa of well established sequence information. Additionally, we also found approx. 30 kDa of preliminary sequence information. These sequences should facilitate to complete the sequence of neurotoxin E. Furthermore it should be used for the identification of posttranslational modifications which are of crucial importance for the biological activity of the protein.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL: Mary Frances Bostian	22b. TELEPHONE (Include Area Code): 301-663-7325; 22c. OFFICE SYMBOL: SGRD-RMI-S

DD Form 1473, JUN 86 *Previous editions are obsolete* SECURITY CLASSIFICATION OF THIS PAGE

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

DFH ✓ Where copyrighted material is quoted, permission has been obtained to use such material.

DFH ✓ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

DFH ✓ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

DFH ✓ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

DFH ✓ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

DFH ✓ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

Donald F Flint 91/12/25

PI - Signature DATE

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

PARTIAL SEQUENCING OF BOTULINUM NEUROTOXIN E

Hanspeter Michel, Paul A. Martino, Nian-Zhou Zhu, Jeff Shabanowitz and Donald F. Hunt
University of Virginia, Chemistry Department, Charlottesville, VA 22901

Neurotoxins of botulinum clostridium are scientifically interesting for two reasons. First, they are extremely toxic. Second, they can be used as models for three important biological phenomena, selective recognition by a target cell, transport through the plasma membrane and toxic activity. All three activities are situated on one polypeptide of approx. 150 kDa. Whereas the complete gene sequences of neurotoxins A and C1 were published very recently only parts of the neurotoxin E sequence is known. By using mass spectrometry, supported by automated Edman degradation we were able to deduce approx. 50 kDa of well established sequence information. Additionally, we also found approx. 30 kDa of preliminary sequence information. These sequences should facilitate to complete the sequence of neurotoxin E. Furthermore it should be used for the identification of posttranslational modifications which are of crucial importance for the biological activity of the protein.

TABLE OF CONTENTS

1. INTRODUCTION

Botulinum neurotoxins, produced in Clostridium botulinum, can be classified into seven types, A,B,C_1,D,E,F and G (1). Botulinum neurotoxins are synthesized as aprox. 150 kDa single chain precursor which is not or only weakly toxic. These precursors are then posttranslationally modified into the highly toxic form (2,3). Two types of posttranslational modifications are described. 1) Proteolytic cleavage (nicking) at a special susceptible position into a heavy (aprox. 100 kDa) and a light (aprox. 50 kDa) chain, which are hold together by at least one disulfide bridge (2). 2) activation by proteases (4-7). Nicking alone has not been found to be responsible for the activation of the protein (8). This is supported by the fact that neurotoxins B and E are not nicked but are activated by proteases (4-7). Whereas the site for the nicking is described to be at a well defined position, little is known about the exact mechanism of the actual activation. Recently a trypsin like protease from Clostridium botulinum type A has been purified and characterized (9). This protease cleaves single chain type A botulinum neurotoxin into the two chain form. Although botulinum neurotoxin E exerts its toxicity as intact single chain protein it can easily be nicked by trypsin as well as Lys-C (10).

Botulinum neurotoxins are multifunctional proteins. Their action as highly toxic substances can be described in three different steps. 1) Selective binding to receptors on the surface of the nerve cell plasma membrane. 2) Transfer of the protein through the plasma membrane into the cytoplasm. 3) Catalytic function in the cytoplasm, which produces nerve cell dysfunction. In analogy to other structurally related toxins, different regions of the protein can be attributed with the different functions. For a review see (11,12). Whereas the light chain is believed to contain the catalytic function, the C-terminus of the heavy chain seems to be responsible for selective binding and the N-terminus for internalization.

To fully understand all aspects of action of botulinum neurotoxins exact knowledge of the primary sequence, posttranslational modifications as well as higher order structures is essential. Until recently only partial sequence information of botulinum neurotoxins were available (13-16). Recently the complete sequence of botulinum neurotoxin A (17) and botulinum neurotoxin C1 (18) have been reported. Together with the complete sequence of botulinum neurotoxin A was published a 273 amino acid residues long piece of the N-terminus of botulinum neurotoxin E. These sequences were derived from the corresponding gene sequence. In this report we present aprox. 50 kDa of the primary sequence of the 150 kDa of neurotoxin E.

2. MATERIAL AND METHODS

Botulinum neurotoxin E and a chymotryptic digest were provided by Dr. James Schmidt. All preparations were assayed for non toxicity previous to sending. Trypsin and Glu-C protease were sequencing grade from Boehringer Mannheim. CNBr was from Aldrich. All solvents for high pressure liquid chromatography were HPLC grade. All other chemicals and solvents were of highest available purity.

Purification of botulinum neurotoxins is described elsewhere (19).

Digestion with trypsin. Aprox. 1 nmol of pyridylethylated neurotoxin E was dissolved in 1 µl formic acid. Water was added to a final volume of 100 µl and the ph adjusted to 8.3 by adding solid Tris base. The digestion was done with 3 µg of trypsin (12h, 37 °C). The mixture was acidified to ph 3 with acetic acid and the generated peptides separated by reverse phase HPLC. Sample in 100 µl was injected onto a narrow bore RP300 (2.1 mm x 10 cm) and eluted with 0 % to 60 % of 0.1 % TFA in H_20 and 0.085 % TFA in Acetonitril respectively.

Cyanogen bromide cleavage. 2 nmoles of pyridylethylated neurotoxin E was dissolved in 100 µl 70 % (v/v) formic acid. The cleavage reaction was done at 37° C for 24 hours with 1 mg of cyanogen bromide. The mixture was then lyophilized to remove solvents and cyanogen bromide. The sample was dissolved in 3 µl of formic acid and diluted to 100 µl with 0.1 % of TFA prior to injection onto a narrow bore BU300 (2.1 x 50 mm)reverse phase column. Peptides were eluted with 0 % to 60 % of 0.1 % TFA (v/v) in H_2O and 0.085 % TFA (v/v) in acetonitril respectively.

Digestion with Glu-C. Peptides were dissolved in 50 mM ammonium bicarbonate buffer to a concentration of 1 - 2 µg/µl. 2 % (w/w) enzyme was added and the digestion done for 16 hours at 37 °C. Separation of peptides was done as described above.

Mass spectrometry. Mass spectra were recorded on either a TSQ-70 triple quadrupole instrument (Finnigan-MAT, San Jose, CA) or a quadrupole Fourier transform instrument (21,23). Operation of these instruments for oligopeptide sequence analysis has been described previously (21-24). Sample ionization and volatilization by particle bombardment on the TSQ-70 instrument were accomplished with a cesium ion gun (Antek, Palo Alto, CA) operated at 6 keV. For ion detection, the conversion dynode of this instrument was operated at 15 keV. Samples for analysis on either instrument were prepared by adding 0.5 to 1 µl of 0.1 % trifluoroacetic acid solution containing 10-100 pmol of peptide(s) to 0.5 µl of a monothioglycerol matrix on a gold-plated stainless-steel probe. Electrospray mass spectra were recorded on the TSQ-70 instrument equipped with the newly developed Finnigan electrospray source. The electrospray needle was operated with a voltage differential of 3-5 kV and a sheath flow of 5 µl/min of a 3/1 mixture of methanol/0.5% acetic acid. Collision activated dissociation experiments were conducted at energies of 20-25 eV for doubly charged ions and 15-18 eF for triply charged ions. Argon at a pressure of 3.5 mtorr was employed as the collission gas. Micro-capillary HPLC experiments were conducted with fused silica columns having an inside diameter of 75 microns and a lenghth

of 75 cm. The last 10 cm of the column was filled with C-18 packing material. Peptides were eluted with a gradient of 0-80% acetic acid (0.5%)/acetonitrile over a 20 min period at a flow of 1-2 µl/min.

Peptide methyl esters. 100-400 pmol of peptide(s) were dried and carboxyl groups esterified with 2 M methanolic HCl. The methanolic HCl was freshly made by dropwise adding of 240 µl of acetyl chloride into 1.5 ml of methanol. After cooling (5-10min) 20 µl of methanolic HCl was added to the peptide(s) and the reaction left at room temperature for 2 hours. After removal of the solvents, the peptides were assayed on mass spectrometer.

Automated Edman degradation. Automated Edman degradation was performed by standard methods on a Model 473 Protein sequencer (Applied Biosystems, Foster City, CA). Analysis of PTH amino acids was done on line with a type 140 A HPLC system. Data recording and analysis was done on a McIntosh IIx computer (Apple Computer, Inc., Cupertino, CA) with the Applied Biosystem software.

3. RESULTS

Digestion of botulinum neurotoxin type E was done with different proteases and with cyanogen bromide. One of the problems to obtain complete digests of pyridylethylated neurotoxin E is its relative insolubility in aqueous solvents. We did choose several ways to circumvent this problem. These include the solubilization of the protein in the presence of 6 M guanidine/HCl, in the presence of SDS and CHAPS, or with concentrated formic acid. As proteases are not normally active under these conditions the solvents had to be adjusted to be compatible with the corresponding protease.

Trypsin digestion. So far the most successful and best characterized method is using formic acid prior to the digestion. In figure 1 is shown the HPLC trace of a digest of pyridylethylated botulinum neurotoxin E with trypsin. For this digest neurotoxin E was first dissolved in a minimal volume of concentrated formic acid. Prior to adding the trypsin the solution was diluted and the pH adjusted to 8.3 with Tris-base. Liquid secondary ion mass spectra were recorded on the TSQ-70 mass spectrometer for the fractions 16 to 47. Table 1 lists the most prominent masses found in each individual fraction. Mass spectra of the individual fractions are shown in appendix A. Every fraction contains between 2 to 5 peptides. Being mixtures of a limited number of peptides these fractions are ideal samples to do the sequencing with the triple quadrupole mass spectrometer.

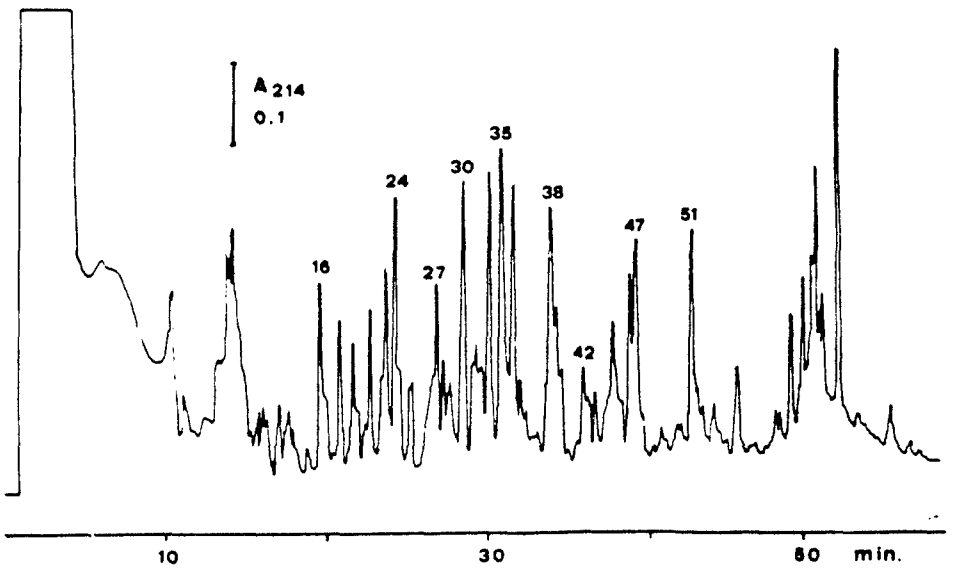


Figure 1. High pressure liquid chromatogram of a tryptic digest of botulinum neurotoxin E. Separation on a reverse phase narrow bore column, RP300 (2.1 x 100 mm).

Table 1: Mass values $(M + H)^+$ of peptides in HPLC fractions of a tryptic digest of neurotoxin E. Masses were recorded by liquid secondary ion mass spectroscopy on the TSQ-70.

fraction	$(M+H)^+$						
16	621	896					
17	779	801	898	1129			
18	895[a]	1131[a]	1227[a]	1376[a]			
19	509	607	886	954	1097		
20	842	886					
21	886	1388[a]					
22	739	1133	1280				
23	750	1086	1117	1134[a]	1262	1569[a]	2263
24	750	1134					
25	750	916	1132	1380			
26	545	608	837	911	1329		
27	1046	1329	2139				
28	1046	1526	1917	1978			
29	784	1292	1526				
30	926	1138	1292	1736	1853[a]		
31	1342[a]	1865[a]	2223				
32	1112	1342	2223				
33	996						
34	1376[a]	1504[a]	1694[a]				
35	947[a]	1152[a]	1779[a]	1876[a]	2802[a]		
36	727	853	1039	1901			
37	1042	1223	1513	1555	2467		
38	755	1436[a]	2409[a]	2470			
39	1264[a]	1420[a]	2470[a]				
40	1089	1266	1458				
41	1365	1719	1820	2287			
42	1715[a]	2308[a]					
43	1157	1244					
44	1157	1604	1969	2513			
45	900[a]	2512[a]					
46	2009	2835					
47	1898[a]	2012[a]					

[a]sequences of peptides determined (see table 2)

Generally the mass of the peptide to be sequenced is selected in the first quadrupole. In the second quadrupole this selected peptide is subjected to fragmentation by collision with argon. Resulting masses of the fragments are analyzed in the third quadrupole. Normally the recording of one collission activated mass spectrum is insufficient for the complete determination of the sequence. Additional information has to be obtained. With the exception of the differentiation of isoleucine and leucine, which do have the same molecular masses this additional information can normally be obtained by subjecting the peptide(s) to selective modification prior to another mass spectral analysis. Whereas esterification in methanolic HCl results in the identification of carboxyl groups, acetylation is normally used to identify free amino groups. We also used automated Edman degradation. Also the combination of mass spectrometry with automated Edman degradation showed to be very favourable under certain conditions. As an example the sequencing of peptides contained in tryptic fraction 35 is described. Shown in figure 2 is the mass spectrum of fraction 35 which contains five peptides with the masses 947, 1152, 1779, 1875 and 2802. We concluded to be able to obtain collision activated spectra by liquid secondary ion mass spectrometry of the single charged ions of the four peptides 947, 1152, 1779, and 1875.

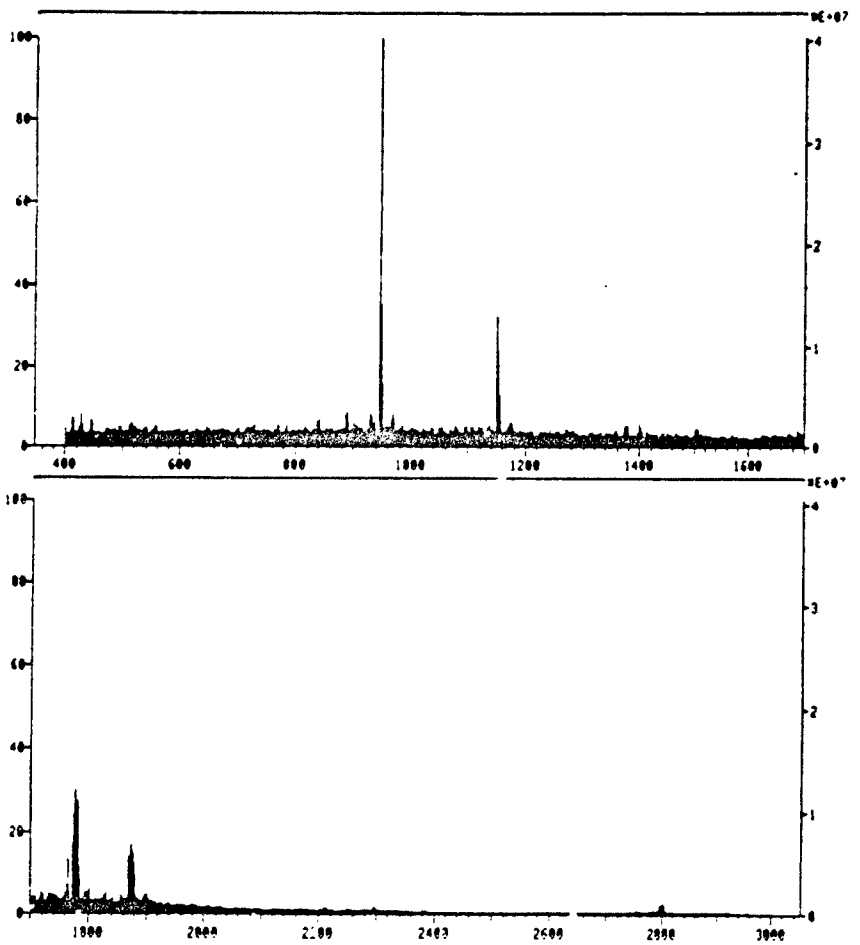
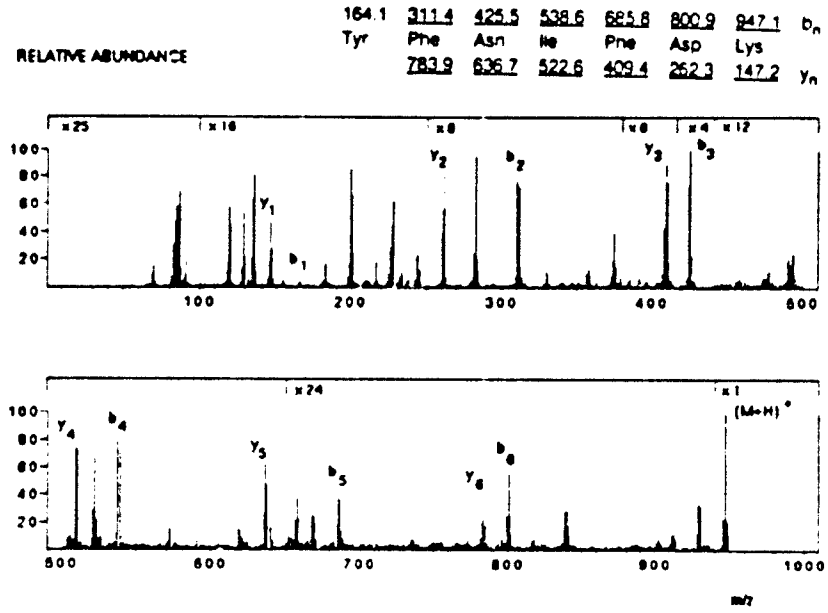
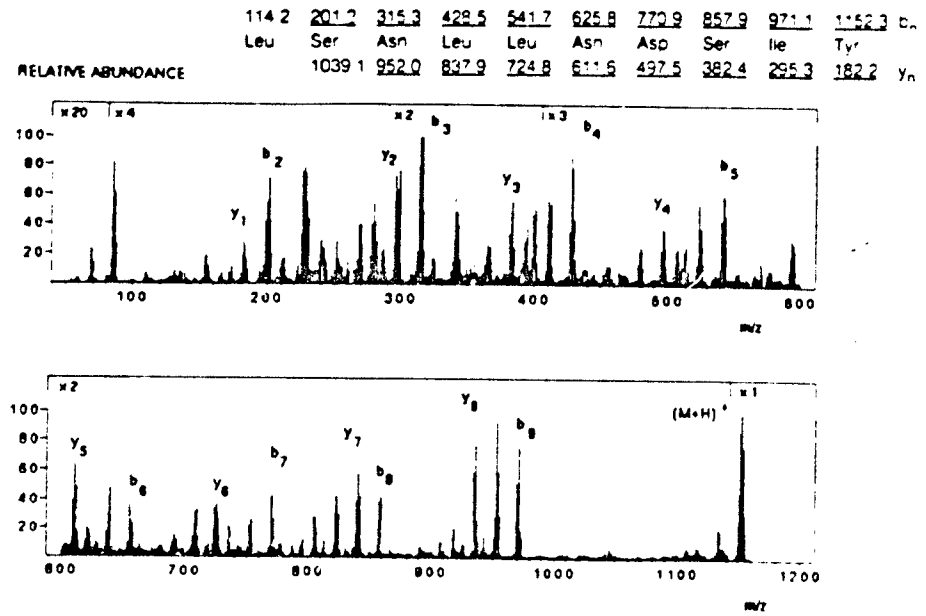


Figure 2. Mass spectrum recorded on HPLC fraction 35 of the tryptic digest of botulinum neurotoxin E.

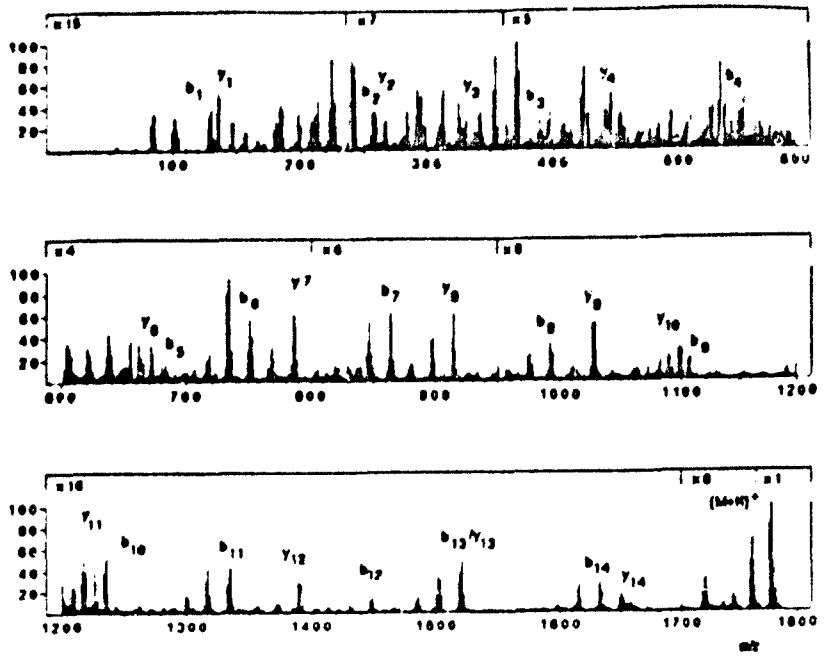
The collision activated mass spectra of these peptides are shown in figure 3a - 3d. For the peptide 2802 we decided to choose electrospray ionization and recorded the spectrum of the triple charged ion. The collision activated mass spectrum is shown in figure 4. Although sequence information can be obtained from all these mass spectra, further information is needed to obtain a complete sequence for all five peptides. We decided to subject the total fraction to automated Edman degradation. The cycles of these degradation are shown in figure 5. Note that no sequence information can be obtained from these cycles due to the complexity of the fraction.





130 1 258.3 369.4 552.6 680.8 751.8 864.9 993.0 1107.1 1235.3 1334.4 1448.5 1519.6 1632.7 1779.0 b_n
Glu Gln Met Tyr Gln Ala Leu Gln Asn Gln Val Asn Ala Ile Lys
1649.9 1521.8 1390.8 1227.4 1099.3 1028.2 915.0 786.9 672.8 544.7 445.5 321.4 250.3 147.2 y_n

C



114 2 185.2 322.- 446.5 609.7 666.8 780.9 851.9 966.0 1023.1 1136.2 1223.3 1338.4 1501.6 1614.8 1728.9 1875.0 b_n
Leu Ala Phe Asn Tyr Gly Asn Ala Asn Gly Ile Ser Asp Tyr Ile Asn Lys
1761.9 1690.8 1543.6 1429.5 1266.4 1209.3 1095.2 1024.1 910.0 853.0 739.8 652.7 537.6 374.5 261.3 147.2 y_n

D

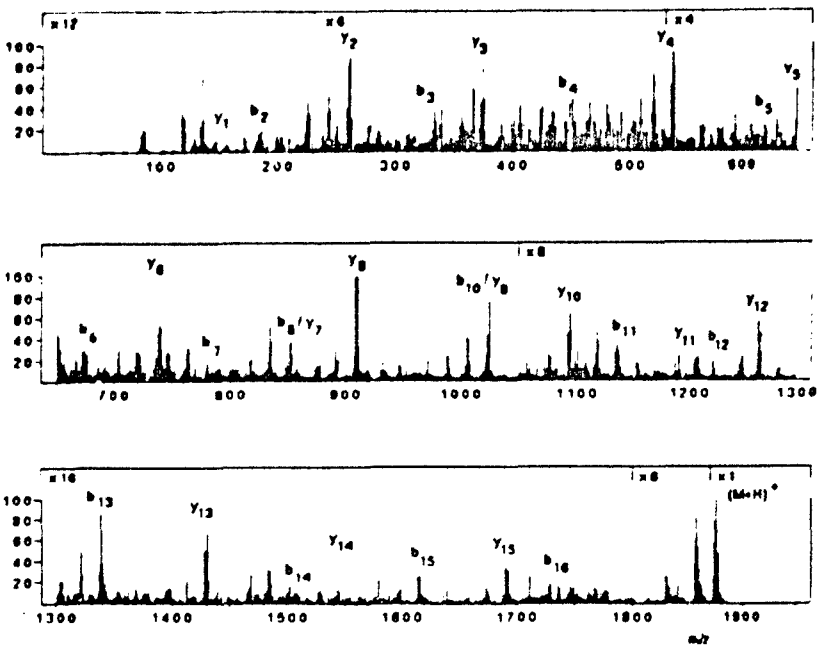


Figure 3. CAD mass spectra of tryptic peptides of botulinum neurotoxin E recorded on the $(M+H)_+$ ions at m/z 947 (a), 1152 (b), 1779 (c) and 1875 (d). Possible fragment masses are indicated on the top. Underlined are fragments which are identified in the mass spectrum.

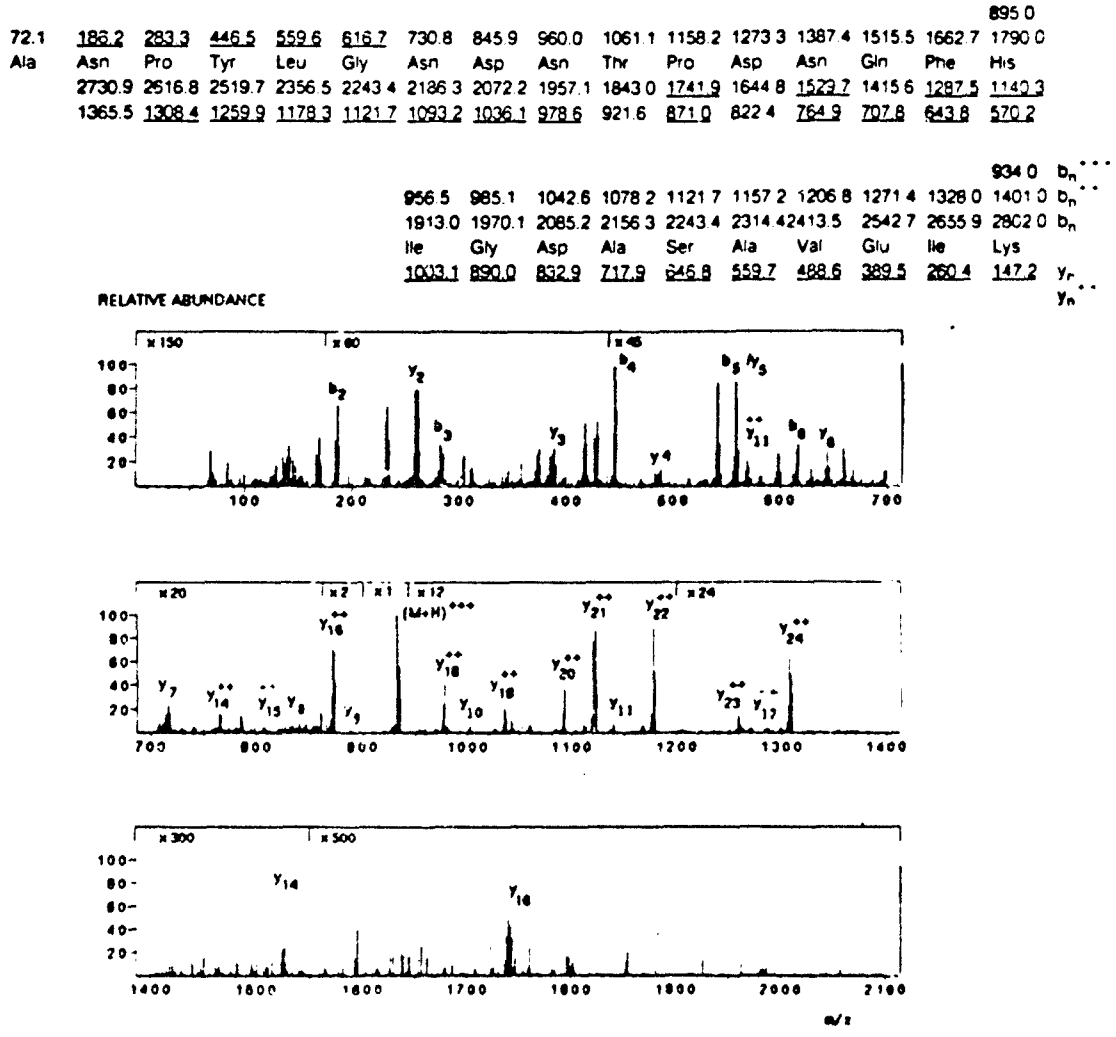


Figure 4. CAD mass spectrum of the tryptic peptide of botulinum neurotoxin E recorded on the $(M+H)^{***}$ ions at m/z 934. Possible fragment masses are indicated on the top. Underlined are fragments which are identified in the mass spectrum.

From the collision activated mass spectra we concluded that the peptides 947, 1779, 1873 and 2802 would contain lysine at the C-terminus, first in all four cases we see the corresponding fragment y-ion (mass = 147), second we used trypsin for the digestion. For peptide 1152 the identity of the C-terminus is not obvious. Lysine as well as arginine can be excluded due to the lack of the corresponding fragment y-ions (mass = 147 or 175 respectively). However we normally observe some chymotryptic activity in the trypsin, especially after prolonged time of digestion, which would give at least some clues for the identification of the C-terminus. From the Edman cycles we found lysine in the position 7, 15, 17 and 26. The assignment of residues 17 to 26 of peptide 2802 is straightforward by comparison of the collision activated mass spectrum (fig. 4) and the Edman cycles (fig. 5). Note that serine in position 21 connot be seen in the Edman degradation, however it can be identified as mass difference between y_6 and y_5 respectively (fig. 4). Position 17 in the automated Edman degradation shows two amino acid residues, isoleucine and lysine. Lysine is the C-terminus of peptide 1875. Isoleucine is in peptide 2802 (mass difference $y_{10} - y_9$ = 113, fig. 4). Position 16 shows two residues as well, asparagine and histidine. Asparagine is

in peptide 1875 (mass difference $y_2 - y_1 = 114$, or $b_{16} - b_{15} = 114$, fig. 3d). Histidine is in peptide 2802 (mass difference $y_{11} - y_{10} = 137$, fig. 4). Note also that due to the presence of histidine y_{11} can also be seen as doubly charged ion. Position 15 in the automated Edman degradation contains three amino acid residues, phenylalanine, isoleucine and lysine. Lysine is the C-terminus of peptide 1779. Isoleucine is in peptide 1875 (mass difference $y_3 - y_2 = 113$, or $b_{15} - b_{14} = 113$, fig. 3d). Phenylalanine is in peptide 2802 (mass difference $y_{12} - y_{11} = 147$, fig. 4). y_{12} as well can be seen as doubly charged ion, again due to the presence of histidine in position 16 of this peptide. In the same way, step by step, the amino acid residues are assigned to the corresponding peptide. This step by step assignment can be done by starting at either end the N-terminus or the C-terminus. Both ways should finally end in identical sequence assignment.

Chymotryptic digest. As a second example we describe the sequencing of a peptide from a chymotryptic digest. The chymotryptic digest of botulinum neurotoxin E was done by Dr. James Schmidt and the fraction provided for analysis. Shown in figure 6 are the collision activated mass spectra of the peptide 1330 and its methyl ester form, peptide 1372. To interpret the spectra fragments containing the N-terminus can be compared. The mass difference between $(M+H)^+$ and b_{11} is 131, this indicates the presence of either leucine or isoleucine on the C-terminus. This is in agreement with the fact that chymotrypsin was used for cleavage. The next fragment, b_{10} is 113 mass units lower than b_{11}, this indicates the presence of another leucine or isoleucine. The mass difference between b_{10} and b_9 is 87. The third residue from the C-terminus is therefore serine. In a similar way residues are identified step by step. The shift of 42 indicates the presence of three carboxyl groups, the C-terminus and two aspartic or glutamic acids. The following sequence information can be obtained from the analysis of the two spectra: XDGNXXDQ/KQ/KSXX, where X is either leucine or isoleucine. Note also that a differentiation between lysine and glutamine is not possible. Acetylation of amino groups and mass spectral analysis could give the additional information needed. We decided however to subject the fraction to automated Edman degradation, this mainly to also differentiate between leucine and isoleucine which are rather abundant in this particular peptide. From these we found the sequence of the chymotryptic peptide 1330 to be: IDGNLIDQKSIL.

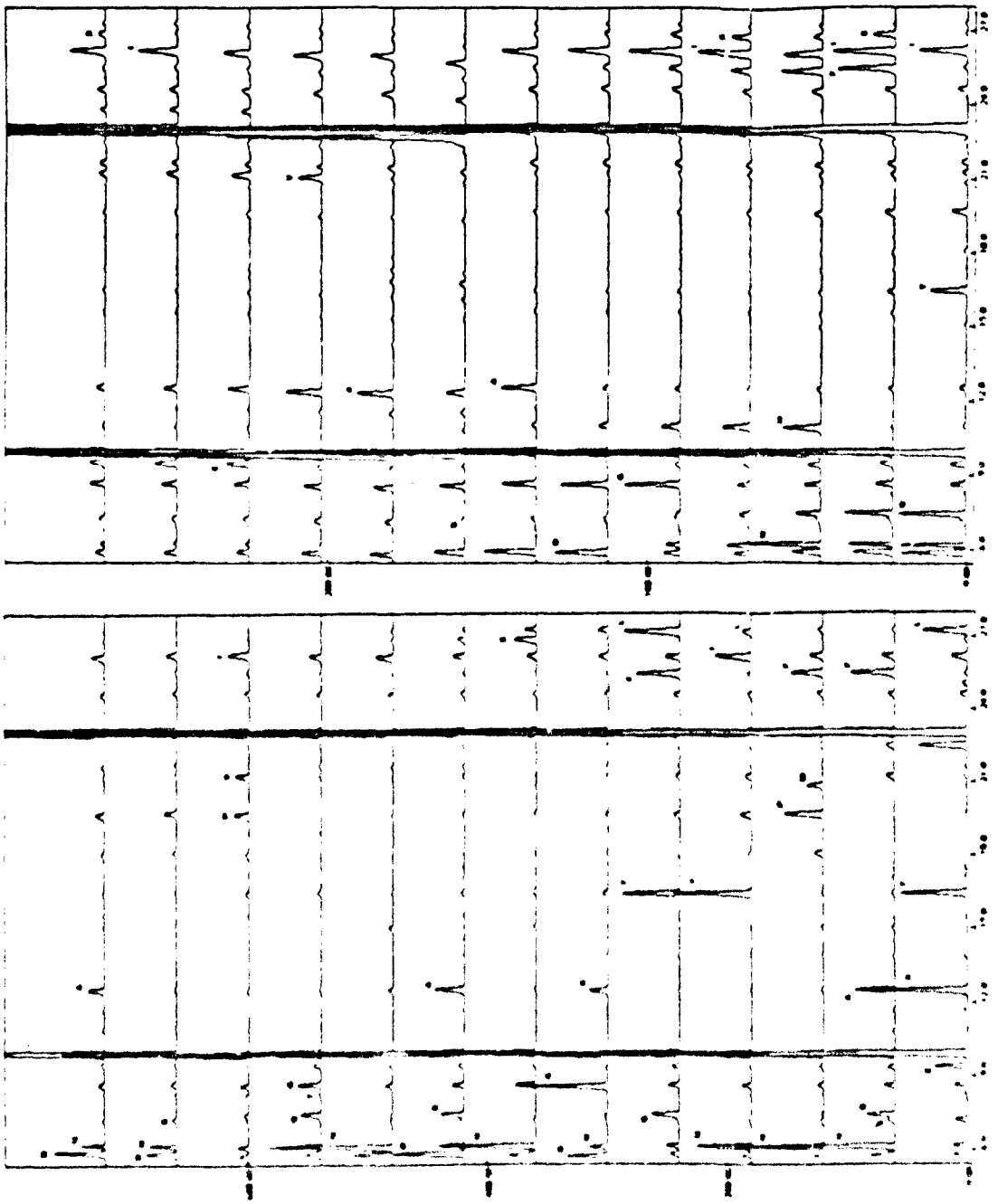


Figure 5. Data obtained from the automated Edman degradation performed on the total HPLC fraction 35 of a tryptic digest of botulinum neurotoxin E. Note the difference in scale between left and right part of the figure.

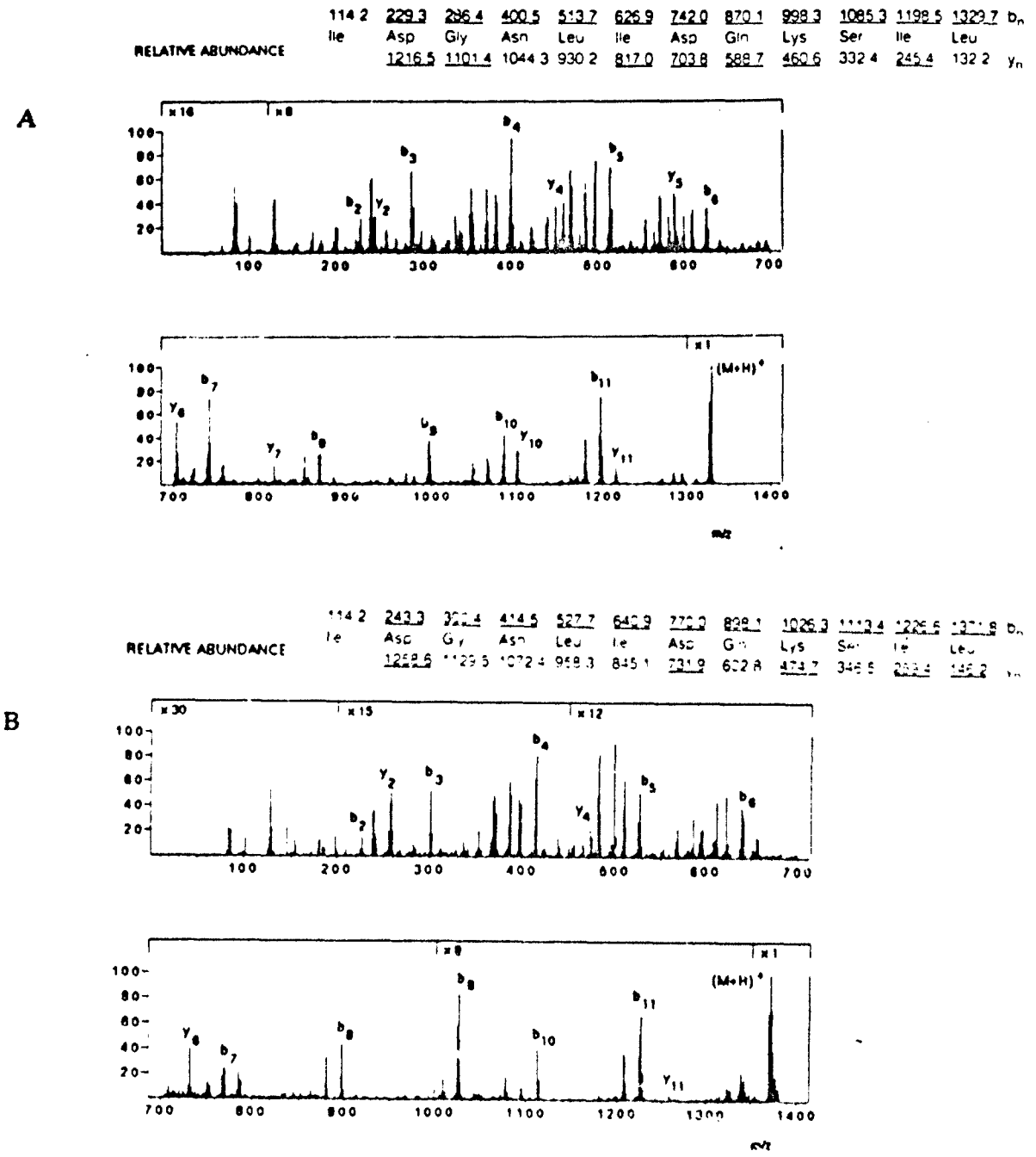


Figure 6. CAD mass spectrum of a chymotryptic peptide of neurotoxin E recorded on the $(M+H)^+$ ions for the unmodified form at m/z 1330 (a) and for the methyl ester form at m/z 1372 (b).

Cyanogen bromide cleavage and Glu-C subdigest. As a last example we present the analysis of a fraction from a cyanogen bromide digestion. Figure 7 shows the HPLC chromatogram of cyanogen bromide treated neurotoxin E. When we subjected the most prominent peak 7 (fig. 7) to automated Edman degradation we found two peptides in this fraction. We were able to obtain information for 22 residues. The following residues eluted in the same stochiometric amount. From this sequencing alone we were not able to deduce any sequence information for the individual peptides.

Residue no.:	1	5	10	15	20
peptide 1:	Y Q A L Q N A V N A I K T I I E N V K T Y L				
peptide 2:	K L I N E V K I R K L R E Y F K A K Y N S I				

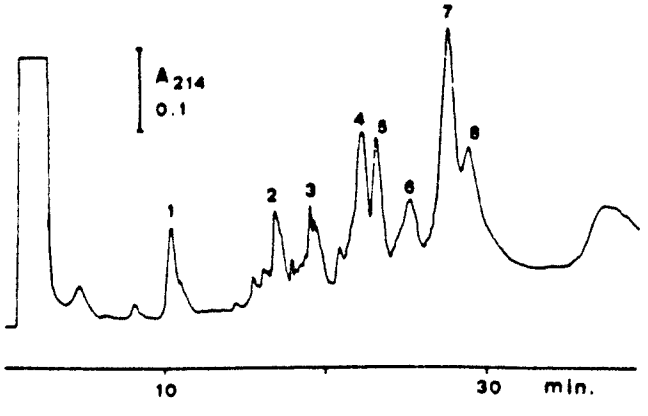


Figure 7. High pressure liquid chromatogram of a cyanogen bromide cleavage of botulinum neurotoxin E. Separation on a reverse phase narrow bore column, BU 300 (2.1 x 50 mm).

To obtain more information from this fraction and to assign the individual amino acid residues from the Edman degradation we subjected this cyanogen bromide fraction to digestion with the protease Glu-C. The HPLC chromatogram of this digest is shown in figure 8.

The individual fractions from this Glu-C digest were then analyzed on the triple quadruple mass spectrometer. CAD mass spectra of the fractions 5 (m/z = 617), 21 (m/z = 1848) and 26 (m/z = 2177) are shown in figure 8. The corresponding sequences are shown on top of the figure with the corresponding fragment masses. All three sequences can be identified as part of the two peptides as found in the automated Edman degradation (see above). From the peptides 2177 and 1848 we can assign the residues to peptide 1. Only one peptide 617 can be assigned to be part of the second peptide in that cyanogen bromide fraction. We have not been able to identify any further peptide. However the partial sequence of this second peptide in cyanogen bromide fraction 7 can be constructed unambiguously.

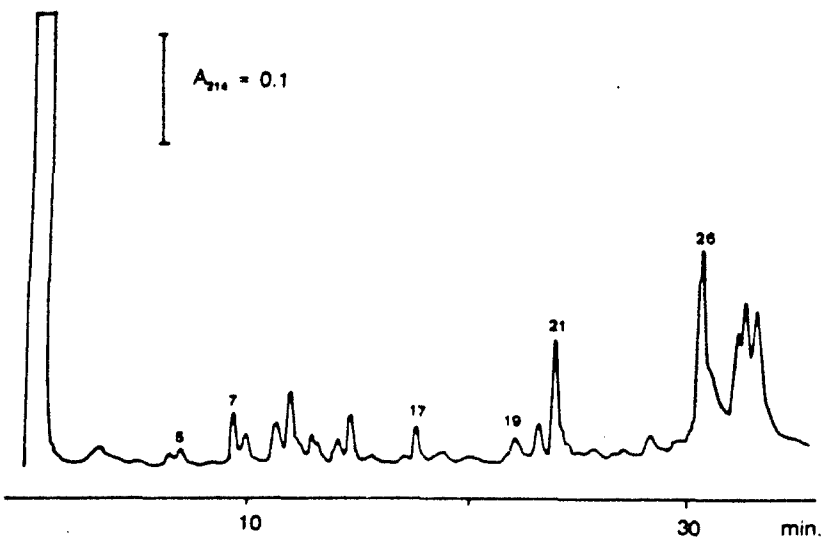
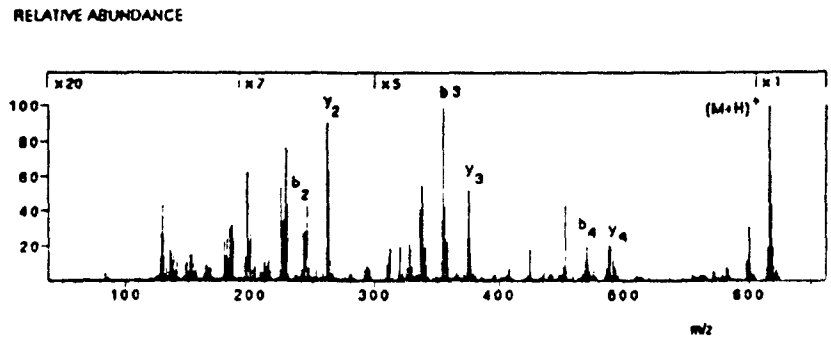


Figure 8. High pressure liquid chromatogram of a subdigest with Glu-C protease. Cyanogen bromide fraction 7 (fig. 7) was subjected to Glu-C (V8) digestion and then separated on a narrow bore reverse phase column, RP 300 (2.1 x 50 mm).

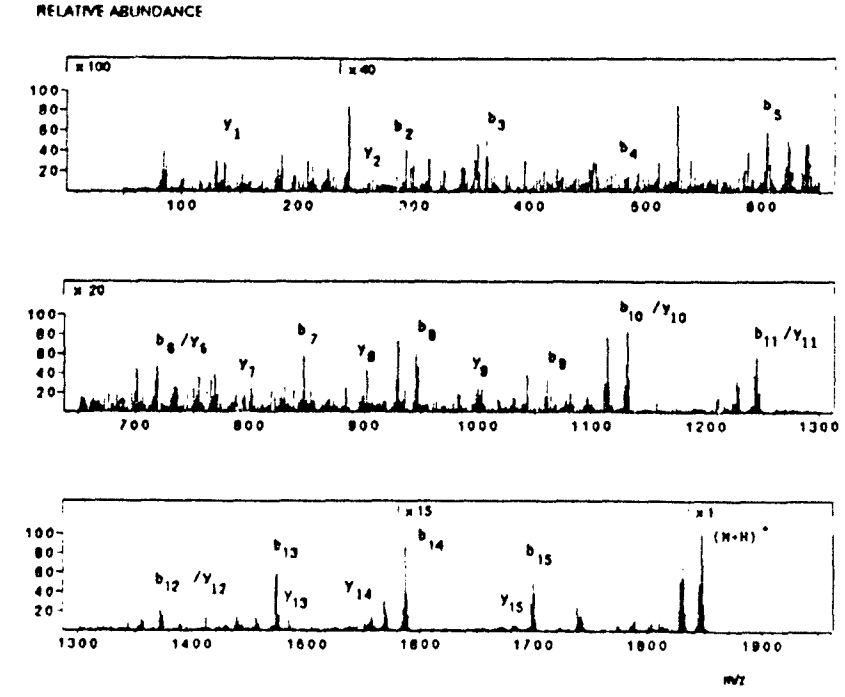
129.2	242.4	355.6	469.7	616.8		b_n
Lys	Leu	Ile	Asn	Glu		
	488.6	375.4	262.2	148.1		y_n

A



164.2	292.3	363.4	476.6	604.7	718.8	846.9	946.0	1060.1	1131.2	1244.4	1372.6	1473.7	1586.9	1700.1	1847.2	b_n
Tyr	Gln	Ala	Leu	Gln	Asn	Gln	Val	Asn	Ala	Ile	Lys	Thr	Ile	Ile	Glu	
	1685.0	1556.9	1485.8	1372.6	1244.5	1138.4	1002.3	903.2	789.1	718.0	604.8	476.6	375.5	262.3	148.1	y_n

B



115.1	214.2	342.4	443.5	606.7	718.9	832.1	947.2	1110.4	1223.6	1336.8	1464.9	1602.0	1659.1	1746.2	1858.4	1972.6	2029.7	2176.8	b_n
Asn	Val	Lys	Thr	Tyr	Leu	Leu	Asn	Tyr	Xxx	Xxx	Gln	His	Gly	Ser	Ile	Leu	Gly	Glu	
	2063.7	1964.6	1836.4	1735.3	1572.1	1458.9	1345.7	1231.6	1068.4	954.2	842.0	713.9	576.8	519.7	432.6	319.4	206.2	148.1	y_n

C

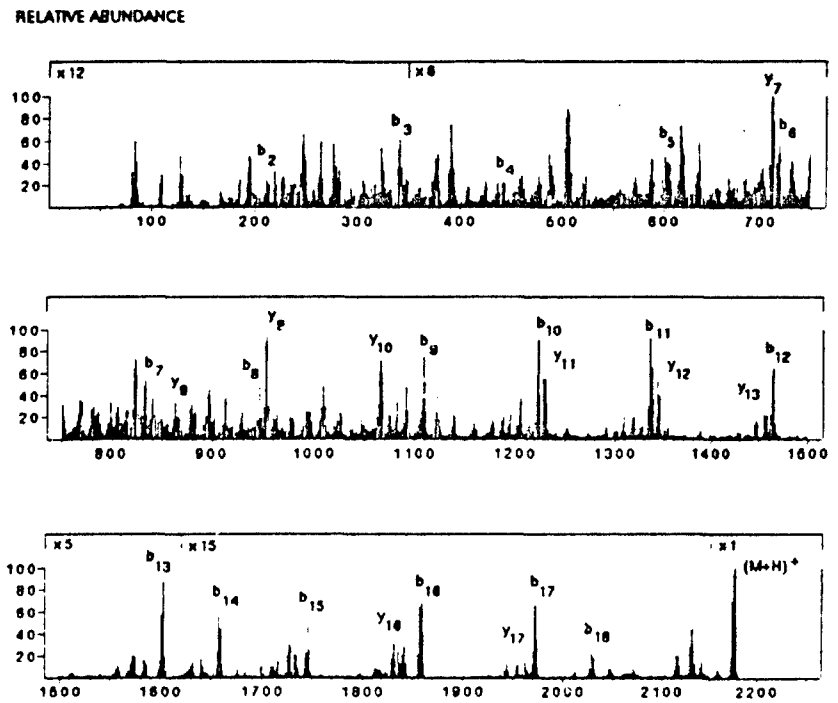


Figure 9. CAD mass spectra of Glu-C protease fragments of botulinum neurotoxin E recorded on the $(M+H)^+$ ions at m/z = 617 (a), m/z = 1848 (b) and m/z = 2177 (c). Fragments are indicated on top. Underlined are fragments which are seen in the mass spectrum.

In a similar fashion we analyzed several other fractions. Table 2 shows a summary of well established sequence information. As a ongoing project a number of fractions are not yet sufficient characterized. More sequence information can be obtained by further detailed analysis of other fractions.

Table 2: Sequences of peptides of botulinum neurotoxin E.

MW[a]	Fraction no.[b]	Sequence[c]
895	Tr-18	SSSVNNMR
1227	Tr-18	QALQNQVNAIK
1376	Tr-18	IKPGGCQEFYK
1388	Tr-21	VQVSNPQLNPYK
1134	Tr-23	VSIAMNNIDR
1569	Tr-23	INSFNYNDPVDNR
2139	Tr-27	YVDTSGYDSNIDINGDVYK
1853	Tr-30	NVIGTTPQDFHPPTSLK
1342	Tr-31	IGLALNIGNEAQK
1865	Tr-31	TILYIKPGGCQEFYK
1376	Tr-34	NNNGNNIGLLGFK
1504	Tr-34	LNLTIQNDAYIPK
1694	Tr-34	THLFPLYADTATTNK
947	Tr-35	YFNIFDK
1152	Tr-35	LSNLLNDSIY
1779	Tr-35	EQMYQALQNQVNAIK
1876	Tr-35	LAFNYGNANGISDYINK
2802	Tr-35	ANPYLGNDNTPDNQFHIGDASAVEIK
1436	Tr-38	LYSFTEFDXATK
2409	Tr-38	VSLNHNEIXWTLQDNAGINQK
1264	Tr-39	WIFVTITNDR
1420	Tr-39	FLTESSISYLMK
2470	Tr-39	VPEGENNVNLTSSIDTALLEQPK
1715	Tr-42	INNNLSGGILLEELSK
2308	Tr-42	VIIMGAEPDLFETNSSNISLR
900	Tr-45	NFSISFW
2512	Tr-45	LSNLLNDSIYNISEGYNINNLK
1898	Tr-47	SILNLGNIHVSNNINFK
2012	Tr-47	EYYLLNVLKPNDFINR
596	Ch-31/15	TIKSF
586	Ch-31/16	MPSNH
749	Ch-45/5	GAEPDLF
1598	Ch-45/6	NYNDPVNDRTILY
964	Ch-47/5	KAINIEEF
1059	Ch-50/2	ENDLQVIL
1330	Ch-49/7	IDGNLIDQKSIL
	CB-21	YQALQNQVNAIKTIIENVKTYLLNYLLQHGSILGESE
	CB-21	KLINEVKIRKLREYDKAKYNSY
	CB-24	.NIWIIPER

2890	CB/V8-5	LSKANPYLGNDNTPDNQFHIGDASAVE
	CB/V8-8	RNVIGTTPQDFHPPTSLK.GDTSY

[a]molecular weight of the $(M + H)^+$ as determined with the TSQ-70 mass spectrometer.
[b]for digestion we used trypsin (Tr), chymotrypsin (Ch), Staph. aureus V8, Glu-C (V8), and cyanogen bromide (CB).
[c]sequences listed in one letter code. X = I or L and period were residue not known.

4. DISCUSSION

Only recently the complete sequences of botulinum neurotoxin A (17) and neurotoxin C1 (18) was presented. Together with the sequence of neurotoxin A was published a apart of the sequence of botulinum neurotoxin E including the N-terminus. Further sequence information of this neurotoxin E could also be expected soon. Homologie alignment of botulinum neurotoxins A and C1 as well as of tetanus toxin is shown in figure 8. The alignment was done with the program CLUSTAL in PCGene (IntelliGenetics Inc., Geneva, Switzerland) which uses the method of Higgins and Sharp (25). In addition the comparison between botulinum toxin A and tetanus toxin has already been described (17). We compared our sequences of botulinum neurotoxin E (table 2) with these three proteins by using the programs SCANSIM and QGSEARCH in PCGene. The region of the highest homology is indicated in figure 10. For most of the peptides (table 2) we found sufficient homology to determine the relative position of the peptide. As can be seen the approx. 40 % of the total possible sequence is distributed very well over the whole range of the protein. This observation is insofar important as it would exclude major parts of the protein from beeing digested and therefore beeing accessible to sequencing. As the tryptic digest (figure 2) is not completely analyzed with regard to sequences, further work has to be done to determine how much of the total sequence can be obtained by analysis of one single digest.

As the number of published gene sequences is increasing the importance of sequence analysis on the level of the protein or the corresponding peptides shifts more towards analysis of posttranslational modifications. Search for such modifications however requires the knowledge of the complete sequence. Neurotoxin E is only partially sequenced to date. Therefore further work is necessary to completely sequence this protein, this can be achieved by sequencing the gene or by continuing the sequencing on the protein level. Once for example the complete sequence is available, very detailed analysis of our data with regard to posttranslational modifications is greatly facilitated. Posttranslational modifications are extremely important for the activation of these group of toxins as already mentioned in the introduction. Up to now very little is actually known about the exact mechanism of activation which means to conversion of the inactive precursor protein to the actual toxic component.

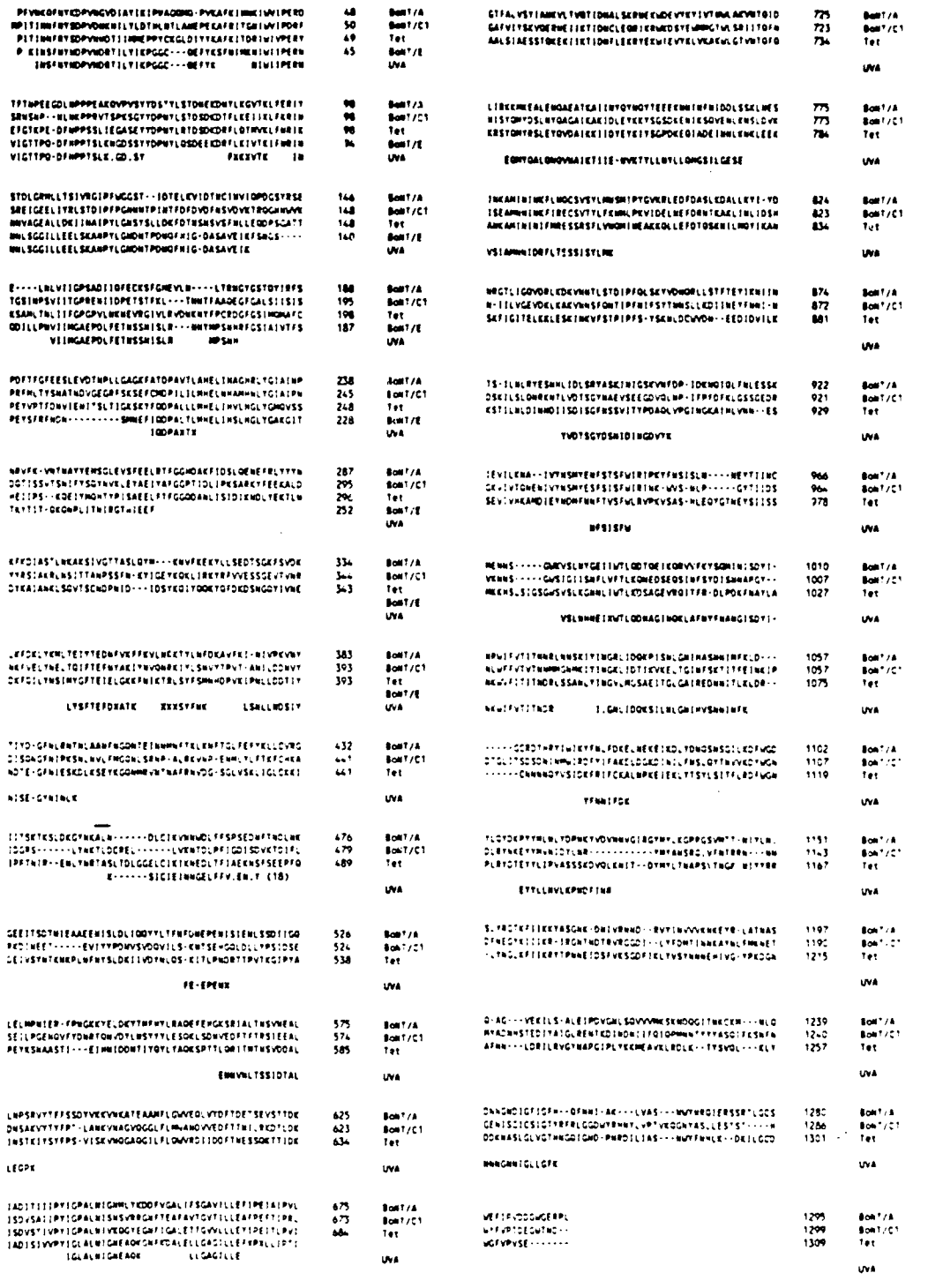


Figure 10. Homology comparison of sequences of botulinum neurotoxins A and C1 and tetanus toxin. And comparison with sequences obtaines in our laboratory (table 2). For sequences see references (13 - 18).

5. REFERENCES

1. Hatheway, C.L., in "Botulinum Neurotoxin and Tetanus Toxin", Simpson, L.L., ed., 1989, Academic press, San Diego, pp. 3-24.

2. DasGupta, B.R., Sugiyama, H., 1972, Biochem. Biophys. Res. Commun., 48, 108-112.

3. DasGupta, B.R., in "Botulinum Neurotoxin and Tetanus Toxin", Simpson, L.L., ed., 1989, Academic press, San Diego, pp. 53-67.

4. DasGupta, B.R., 1971, J. Bacteriol., 108, 1051-1057.

5. DasGupta, B.R., 1972, Biochim. Biophys. Acta, 268, 719-729.

6. Ohishi, I., Okada, T., and Sakaguchi, G., Jpn. J. Med. Sci. Biol., 28, 157-164.

7. Ohishi, I., and Sakaguchi, G., Jpn. J. Med. Sci. Biol., 30, 179-190.

8. DasGupta, B.R., in "Biomedical aspects of botulism", Lewis, G.E., ed., 1981, Academic Press, Inc., New York, pp. 1-19.

9. Dekleva, M.L., and DasGupta, B.R., 1990, J. Bacteriol., 172, 2498-2503.

10. Gimenez, J.A., DasGupta, B.R., 1990, Biochim., 72, 213-217.

11. Simpson, L.L., 1988, Int. Rev. Neurobiol., 30, 123-147.

12. Simpson, L.L., 1986, Ann. Rev. Pharmacol. Toxicol., 26, 427-53.

13. Eisel, U., Jarausch, W., Goretzki, K., Henschen, A., Engels, J., Weller, U., Hudel, M., Habermann, E., and Niemann, H., 1986, EMBO J., 5, 2495-2502.

14. Fujii, N., Oguma, K., Yokosawa, N., Kimura, K., and Tsuzuki, K., 1988, Appl. Environ. Microbiol., 54, 69-73.

15. Gimenez, J., Foley, J., and DasGupta, B.R., 1988, FASEB J., 2, A1750.

16. Wadsworth, C.L., Niece, R.L., Gimenez, J.A., DasGupta, B.R., in "Current Research in Protein Chemistry: Techniques, Structure, and Function", Villafranca, J.J., ed., 1990, Academic Press Inc., New York, pp. 167-178.

17. Binz, T., Kurazono, H., Wille, M., Frevert, J., Wernars, K., and Niemann H., 1990, J. Biol. Chem., 265, 9153-9158.

18. Kimura, K., Fujii, N., Tsuzuki, K., Murakami, T., Indoh, T., Yokosawa, N., Takeshi, K., Syuto, B., and Oguma, K., 1990, Biochem. Biophys. Res. Comm., 171, 1304-1311

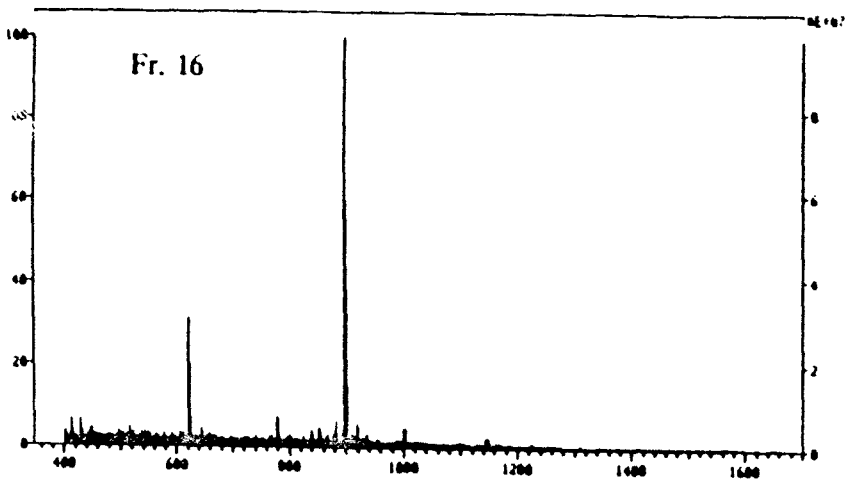
19. Simpson, L.L., Schmidt, J.J., and Middlebrook, J.L., 1988, Meth. Enzymol., 165, 76-85.

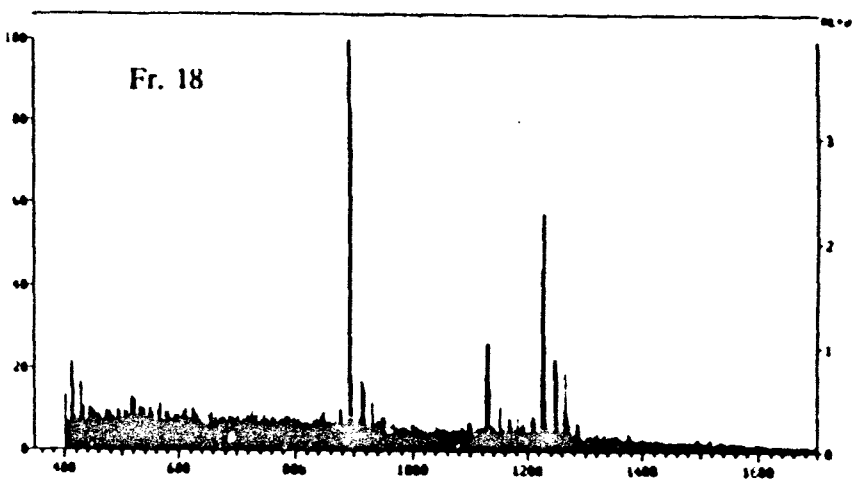
20. Hunt, D.F., Yates, J.R., III, Shabanowitz, J., Winston, S., and Hauer, C.R., 1986, Proc. Natl. Acad. Sci., USA, 83, 6233-6237.
21. Erickson, A.K., Payne, D.M., Martino, P.A., Rossomando, A.J., Shabanowitz, J., Weber, M.J., Hunt, D.F., and Sturgill, T.W., 1990, J. Biol. Chem., 265, (in press)

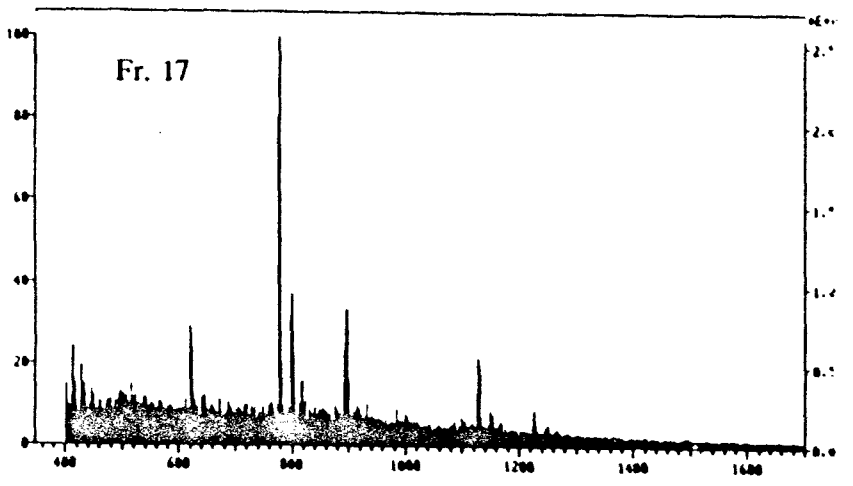
22. Hunt, D.F., Shabanowitz, J., Yates, J.R., III, Zhu, N.Z., Russell, D.H., and Castro, M.E., 1987, Proc. Natl. Acad. Sci., USA, 84, 620-623.

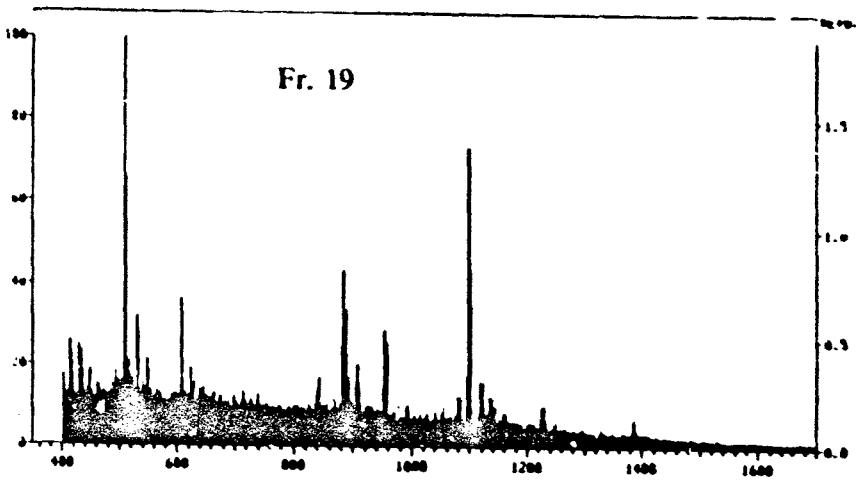
23. Hunt, D.F., Shabanowitz, J., Griffin, P.R., Yates, J.R., III, Martino, P.A., and McCormack, A.L., 1990, in "Proceedings of the Second International Symposium on Mass Spectrometry in the Health and Life Sciences, Burlingame, A.L., and McCloskey, J.A., eds., Elsevier, Amsterdam, in press

24. Higgins, D.G., and Sharp, P.M., 1989, CABIOS, 5, 151-153



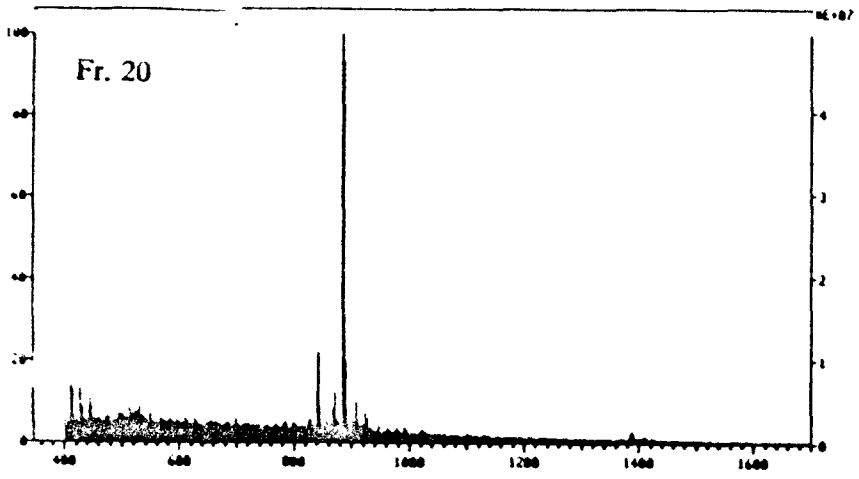


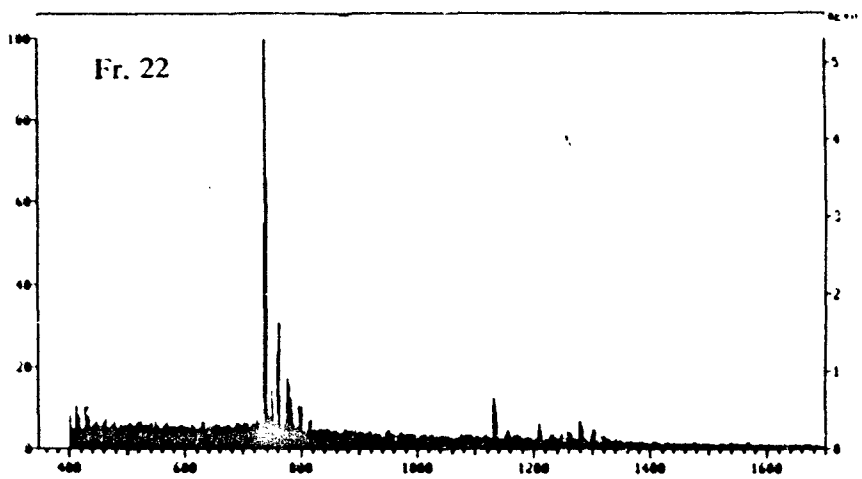


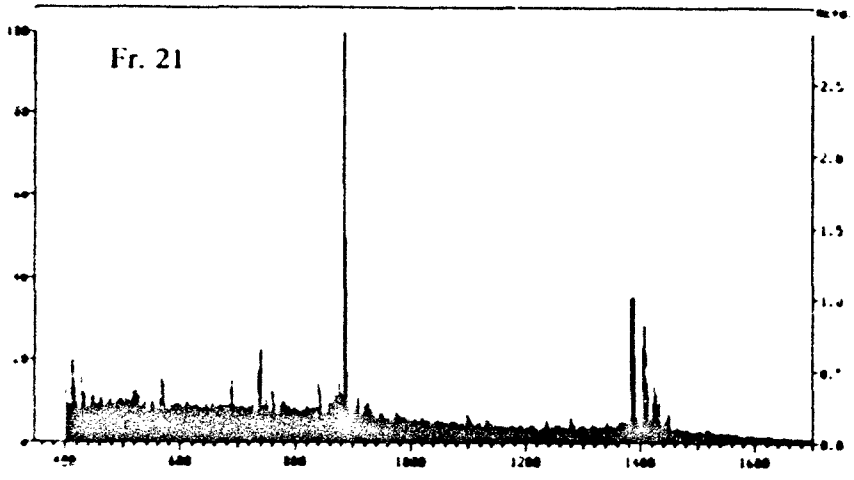


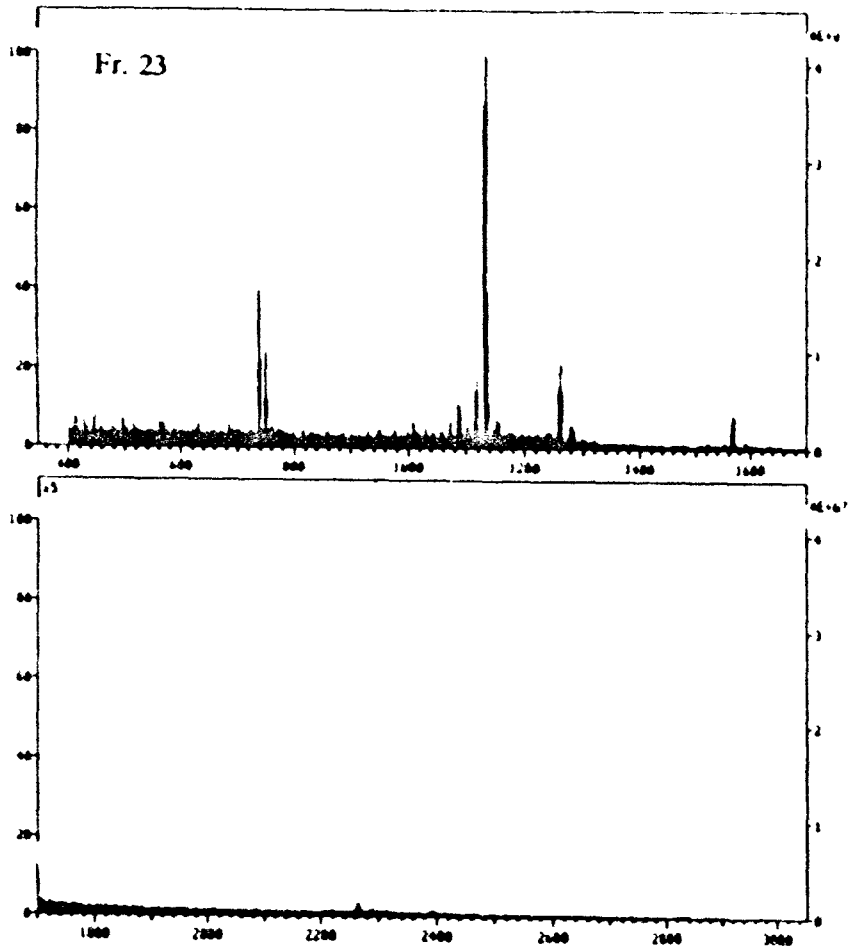
Appendix A

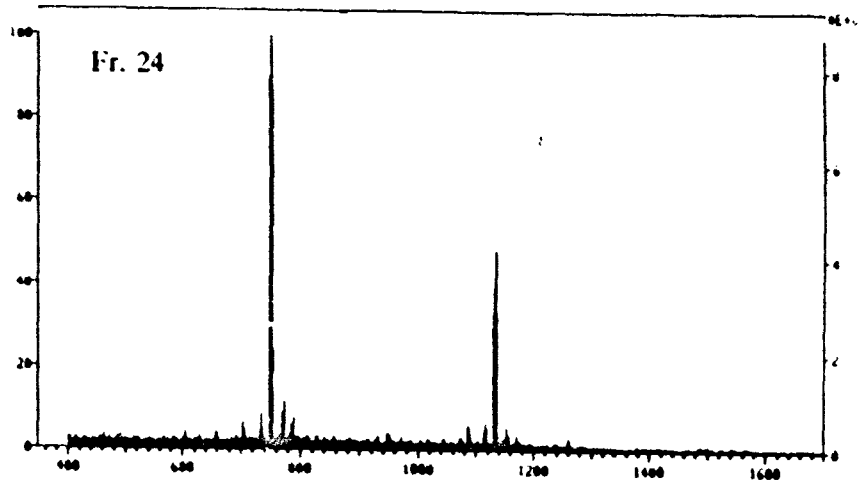
Mass spectra of tryptic fractions 16 to 47 of botulinum neurotoxin E.


Fr. 20

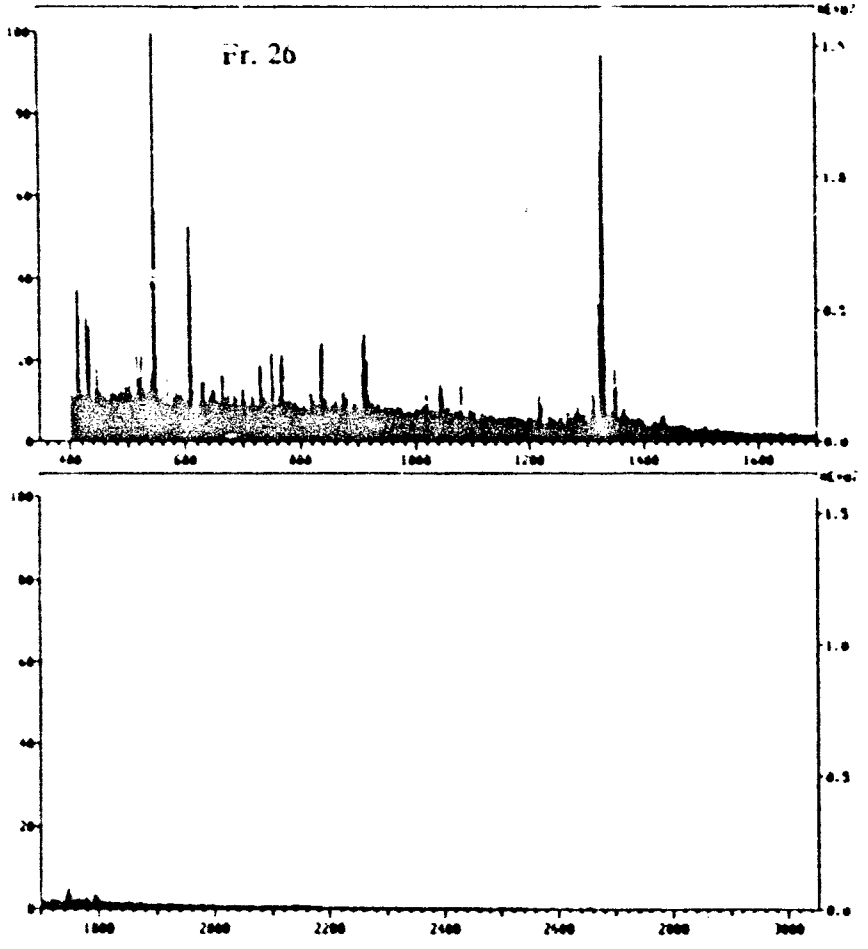

Fr. 22

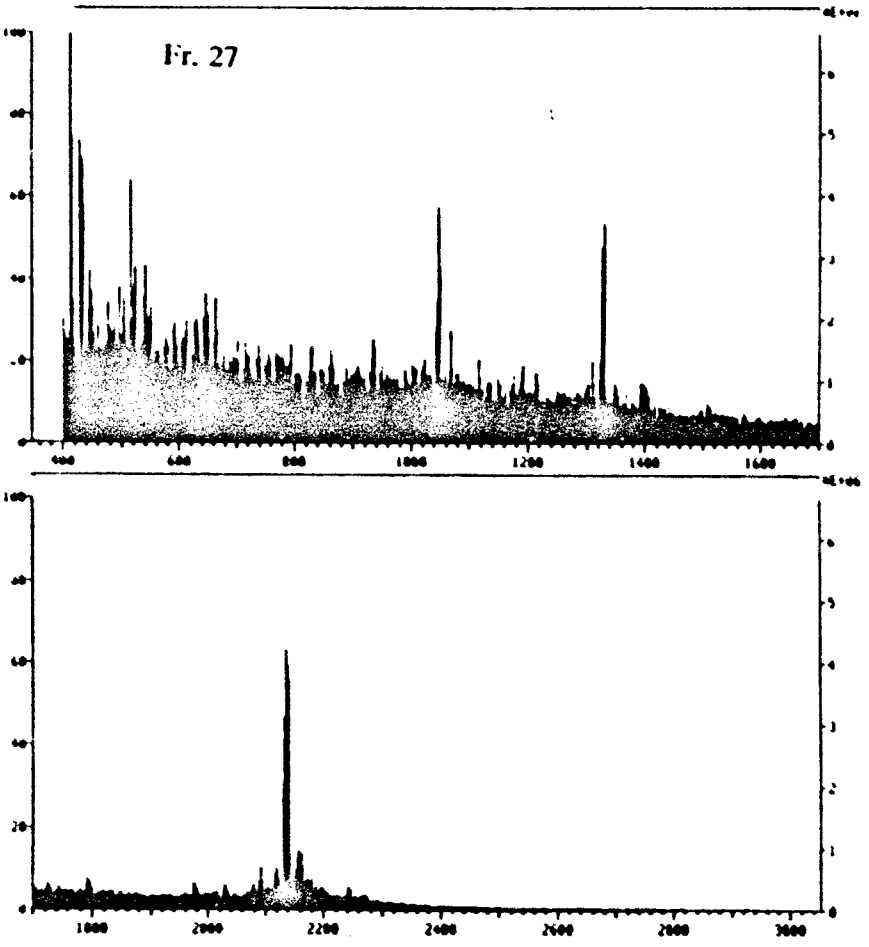

Fr. 21

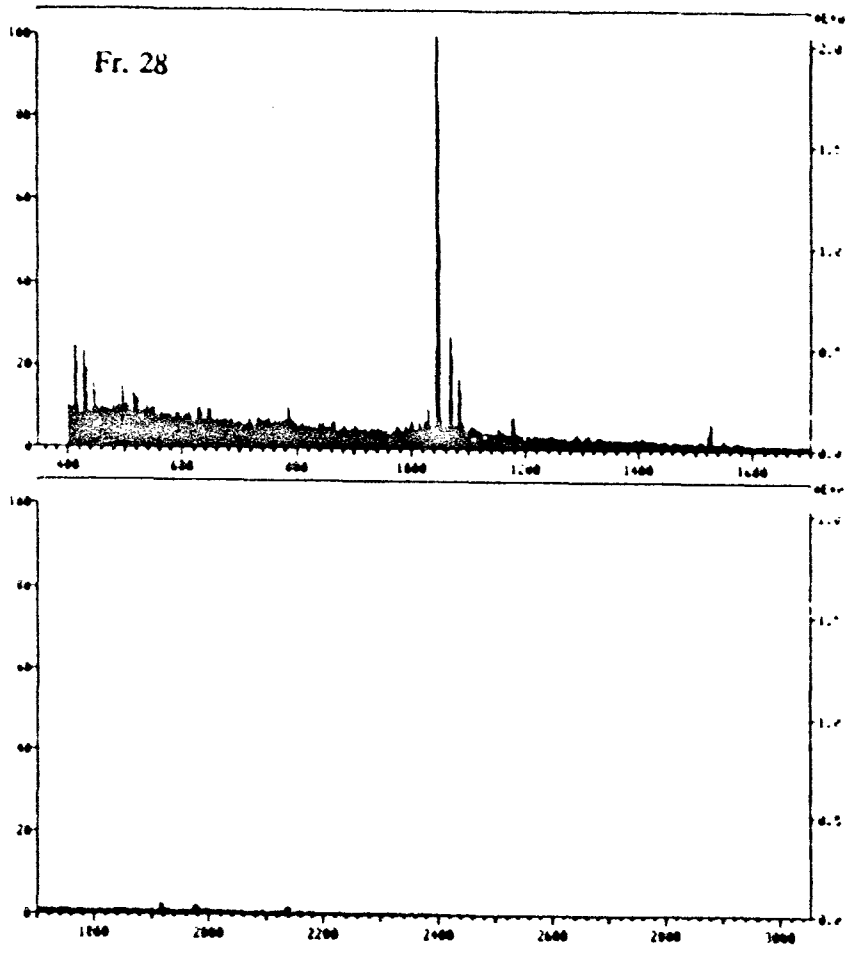

Fr. 23

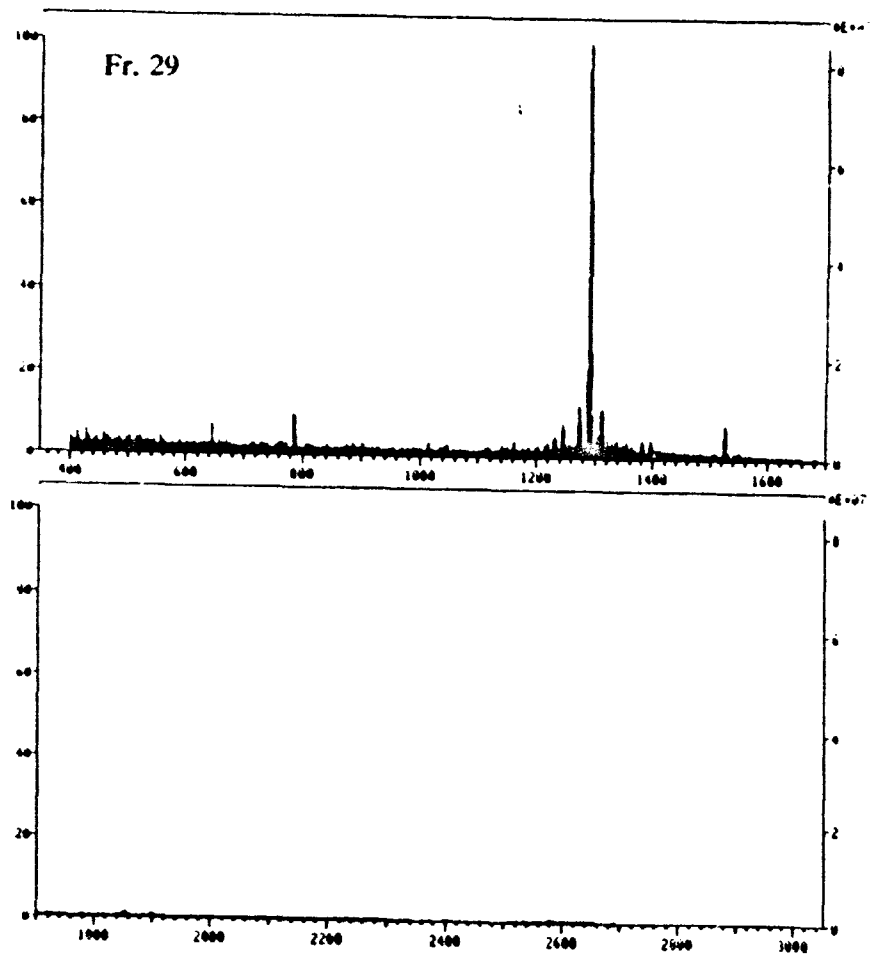

Fr. 24

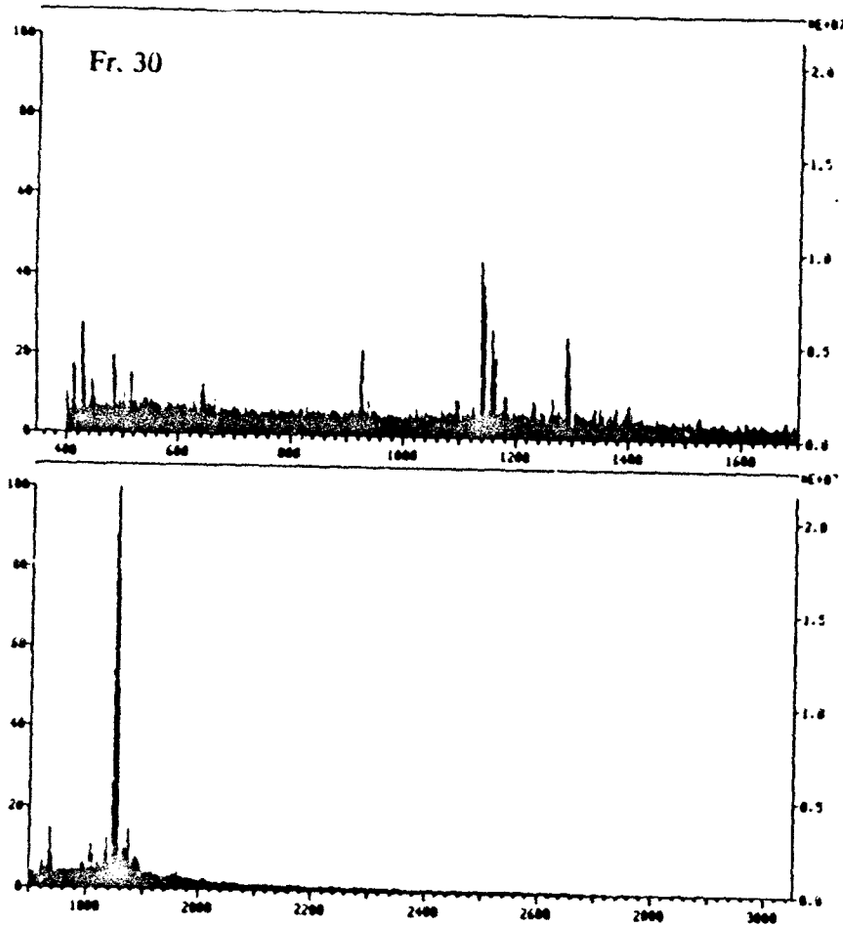

Fr. 25

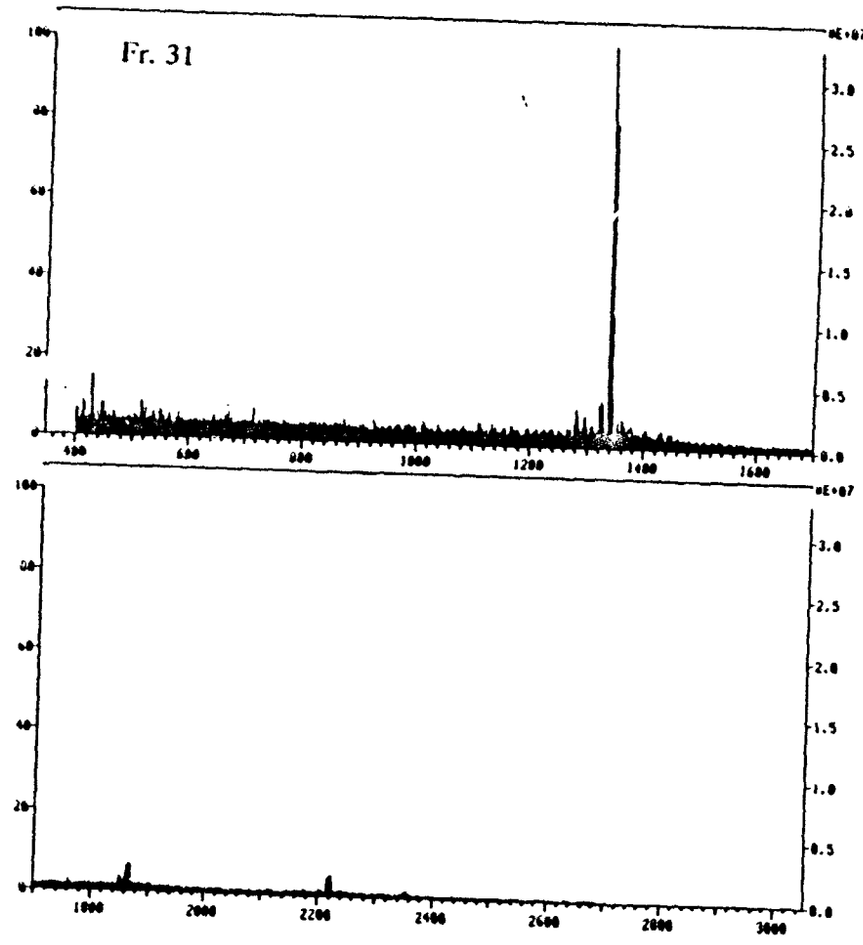

Fr. 26

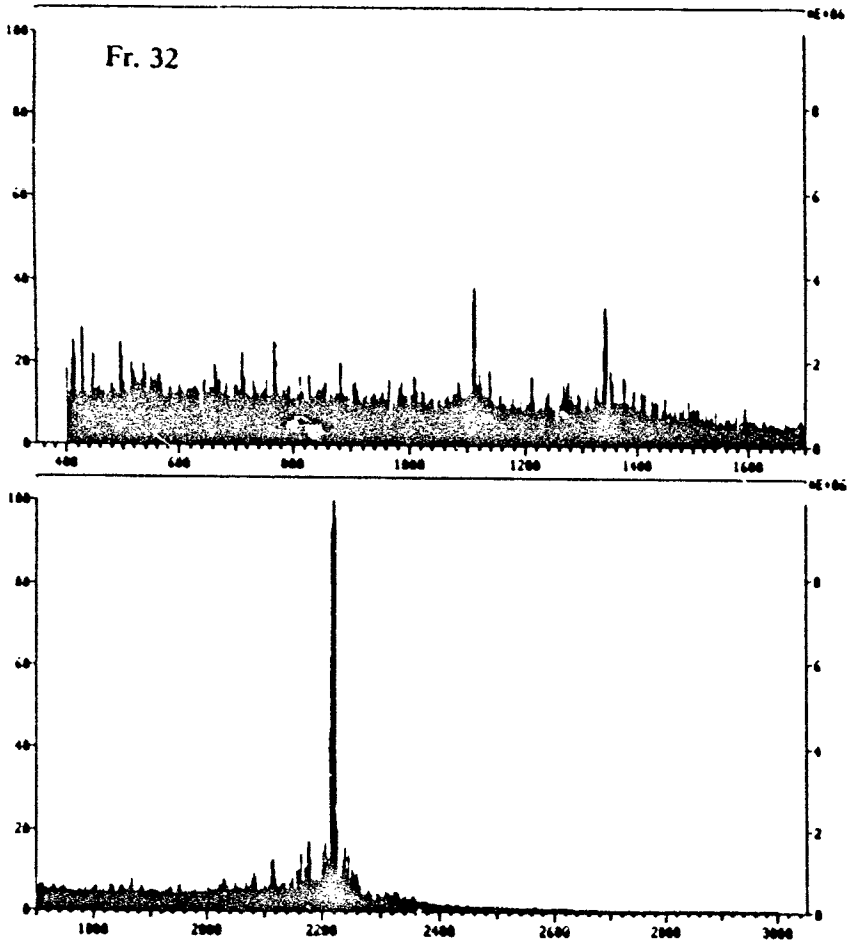

Fr. 27

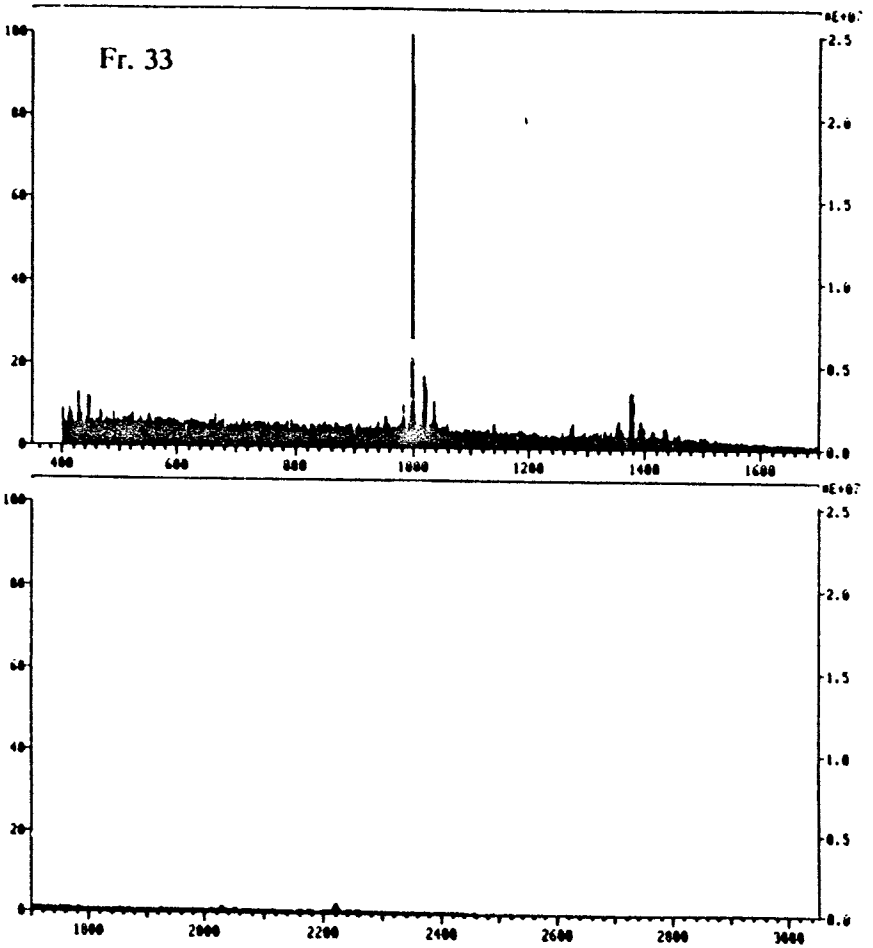

Fr. 28

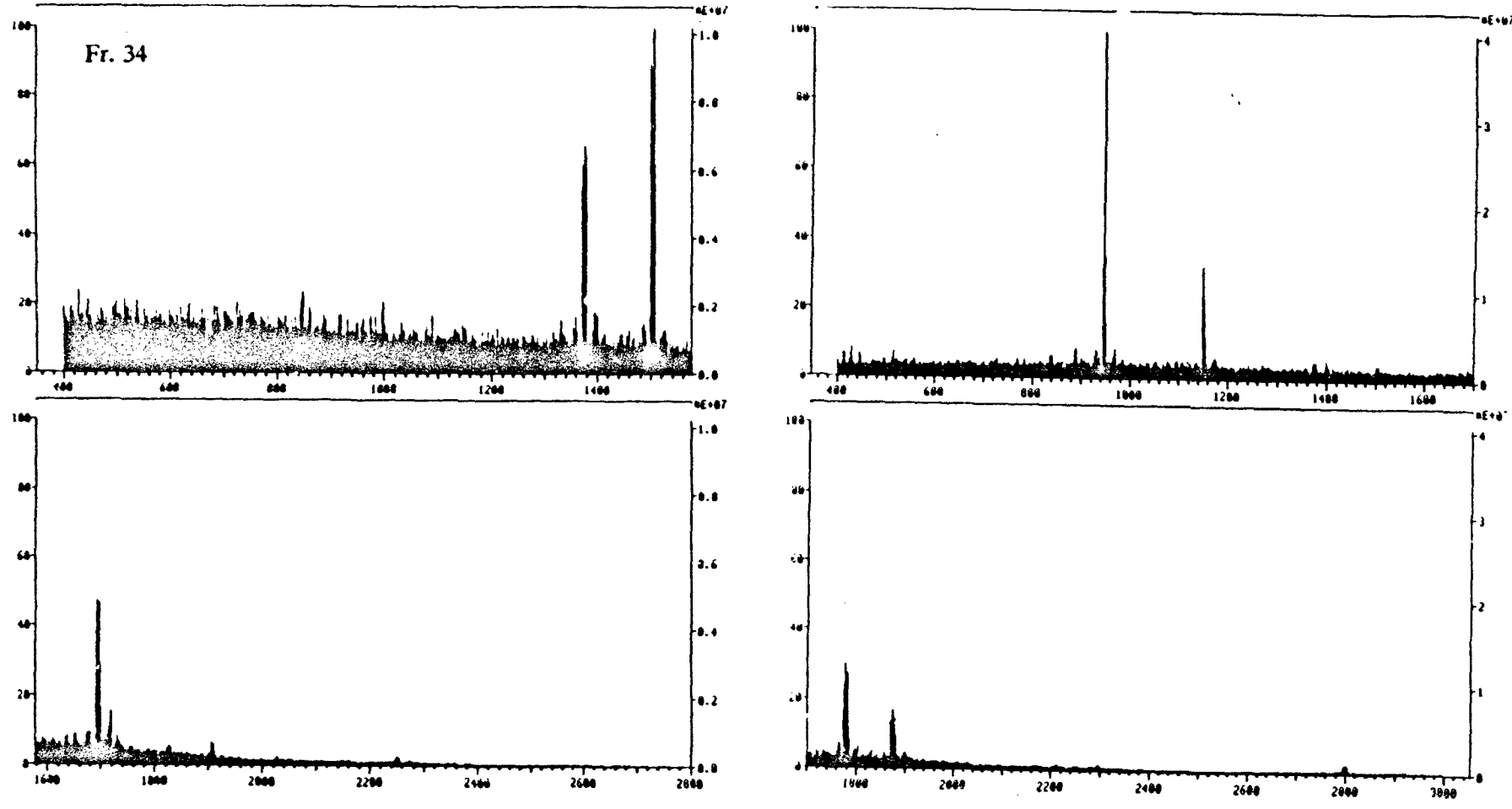

Fr. 29

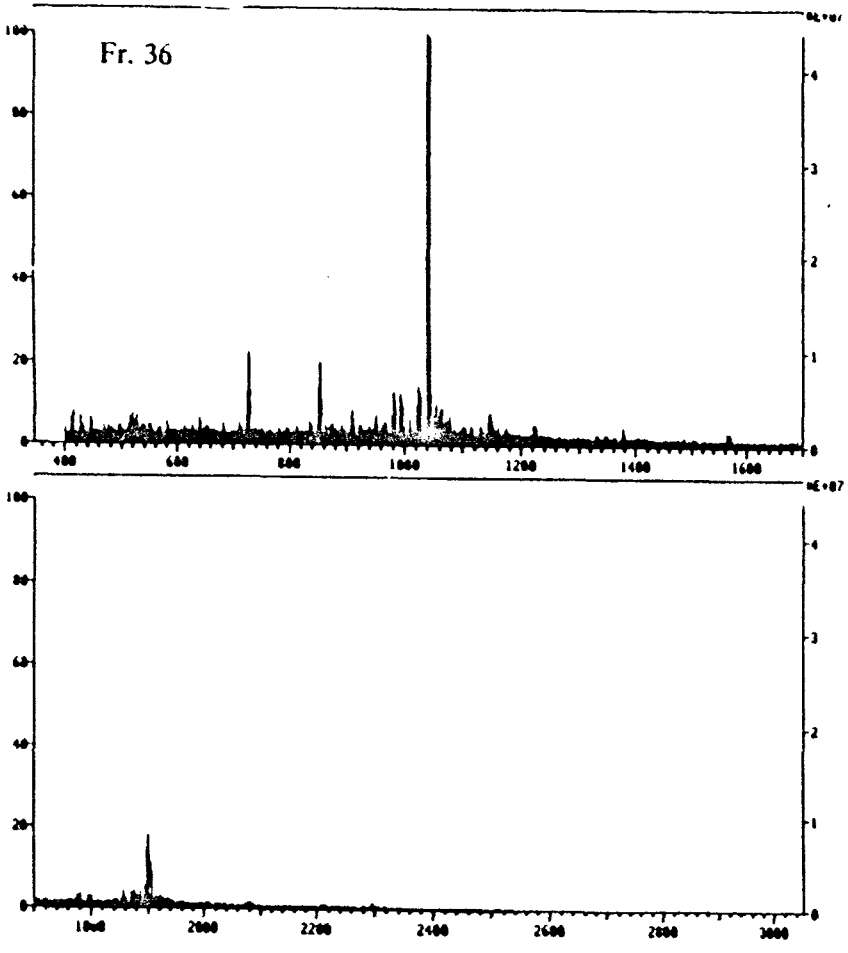

Fr. 30

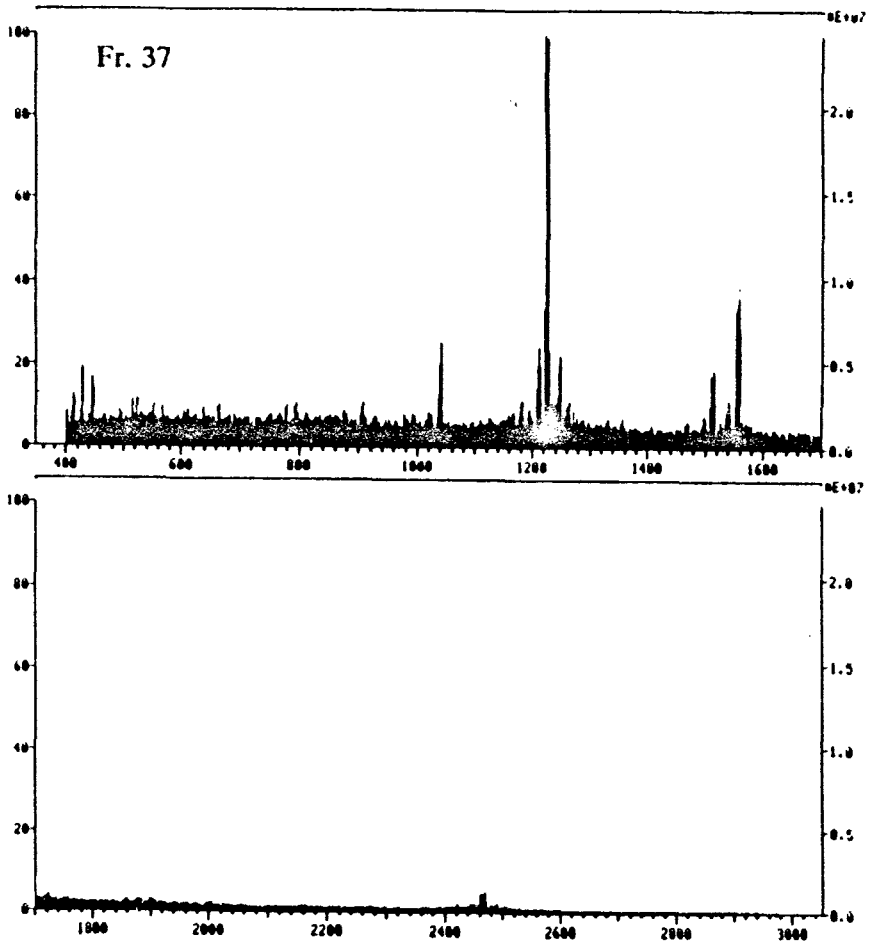

Fr. 31

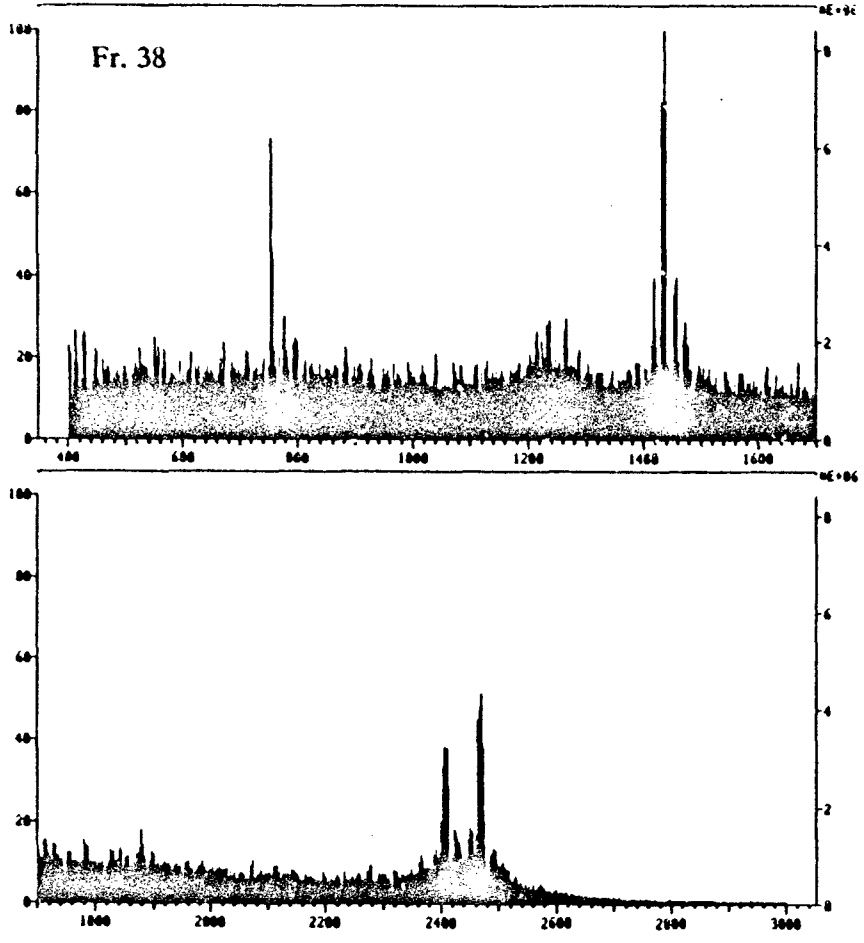

Fr. 32

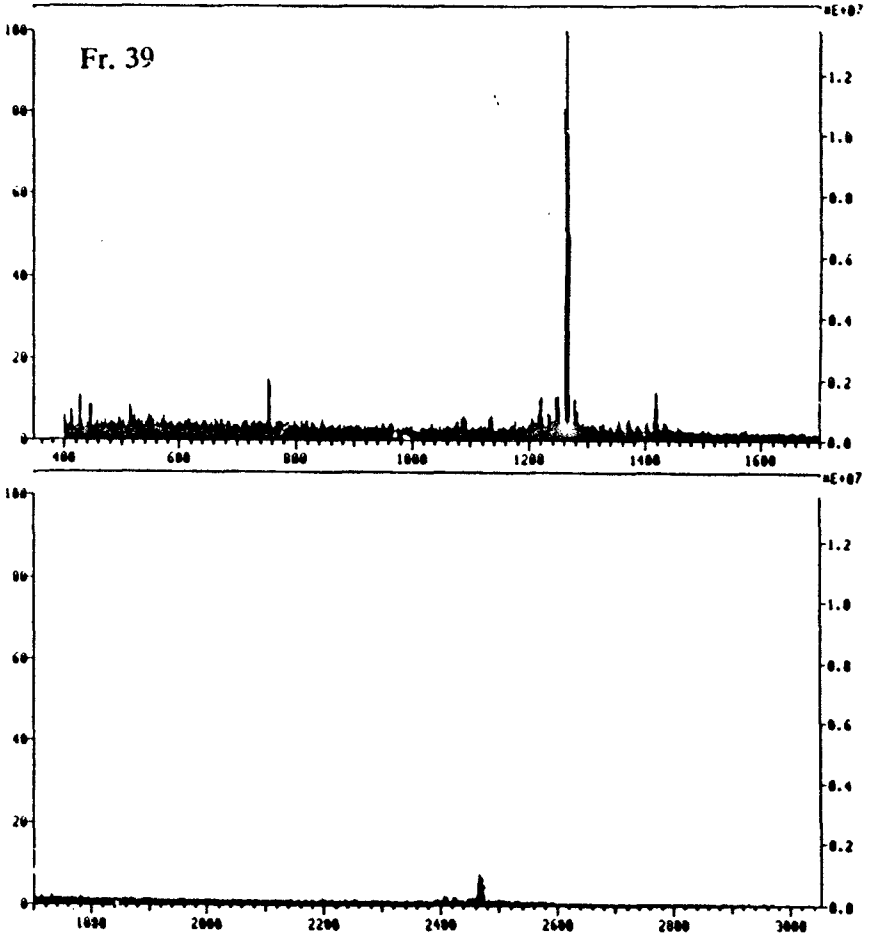

Fr. 33

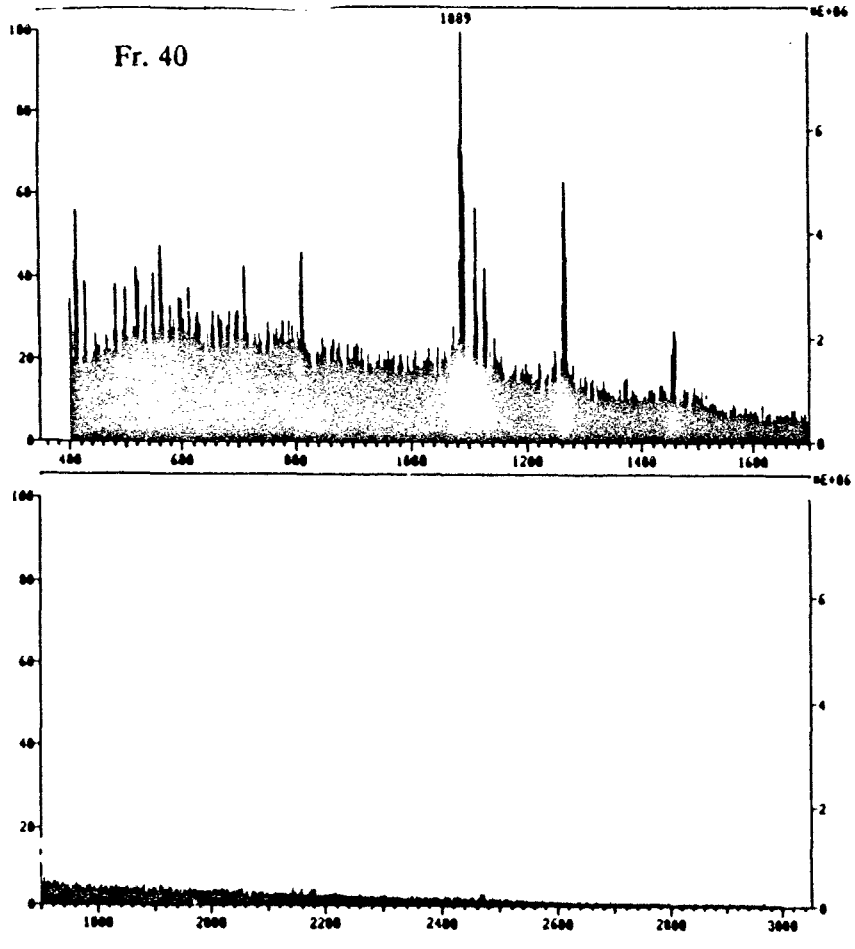

Fr. 34

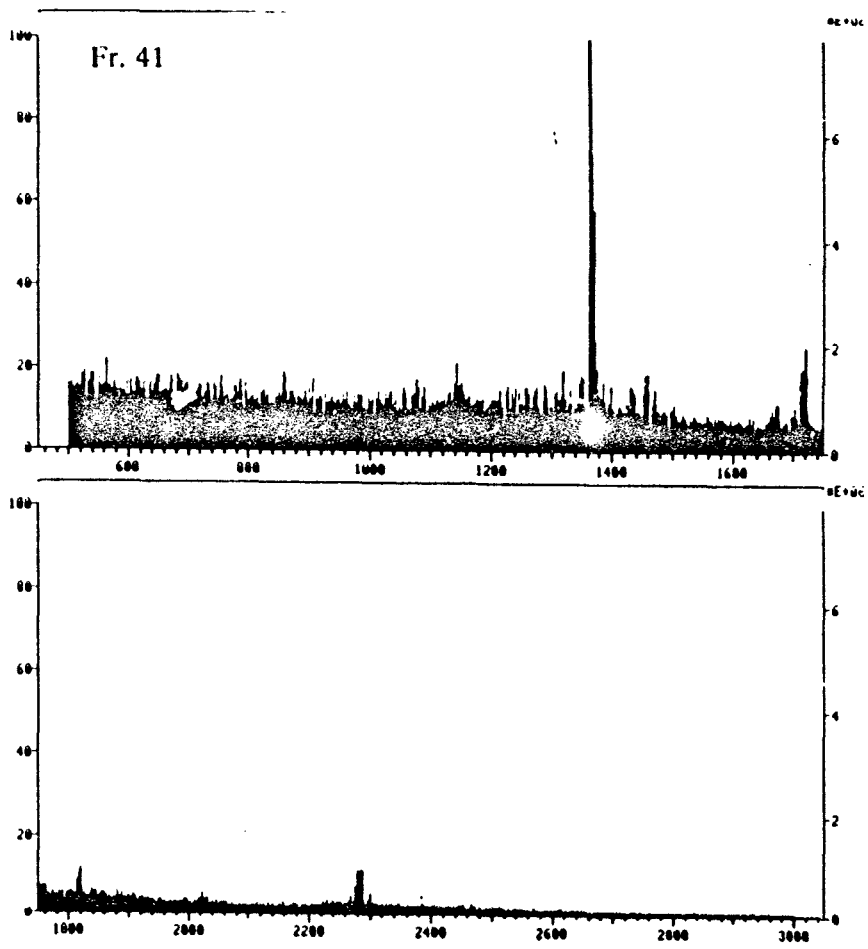

Fr. 36

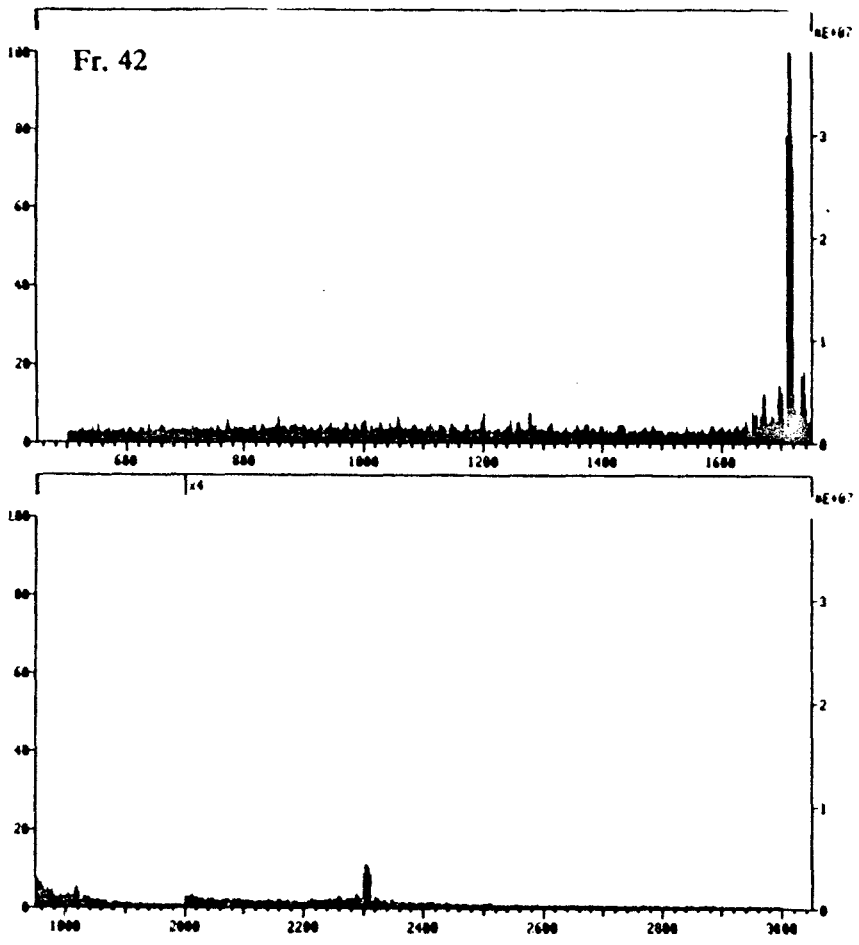

Fr. 37

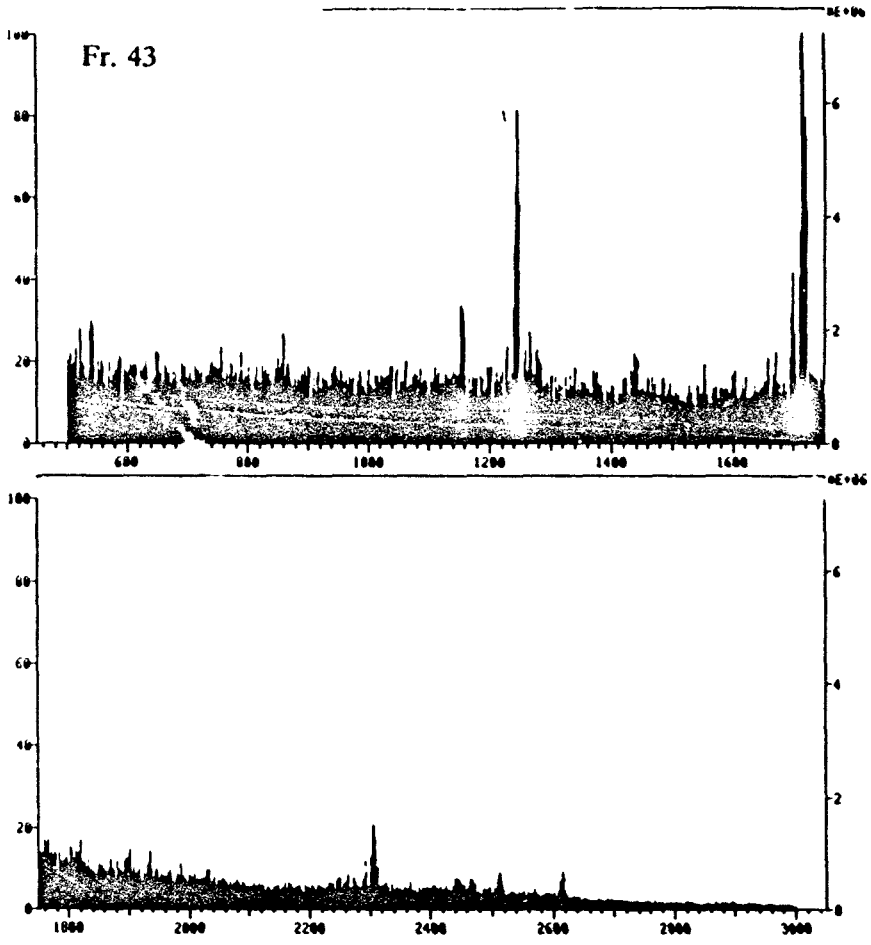

Fr. 38

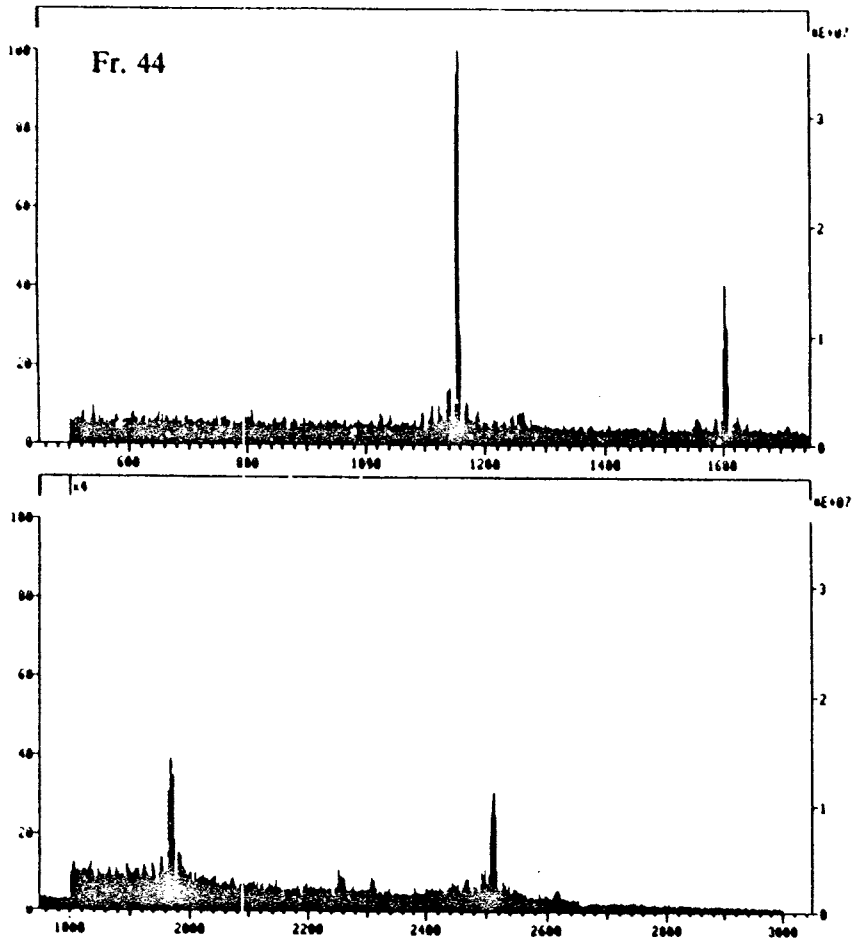

Fr. 39

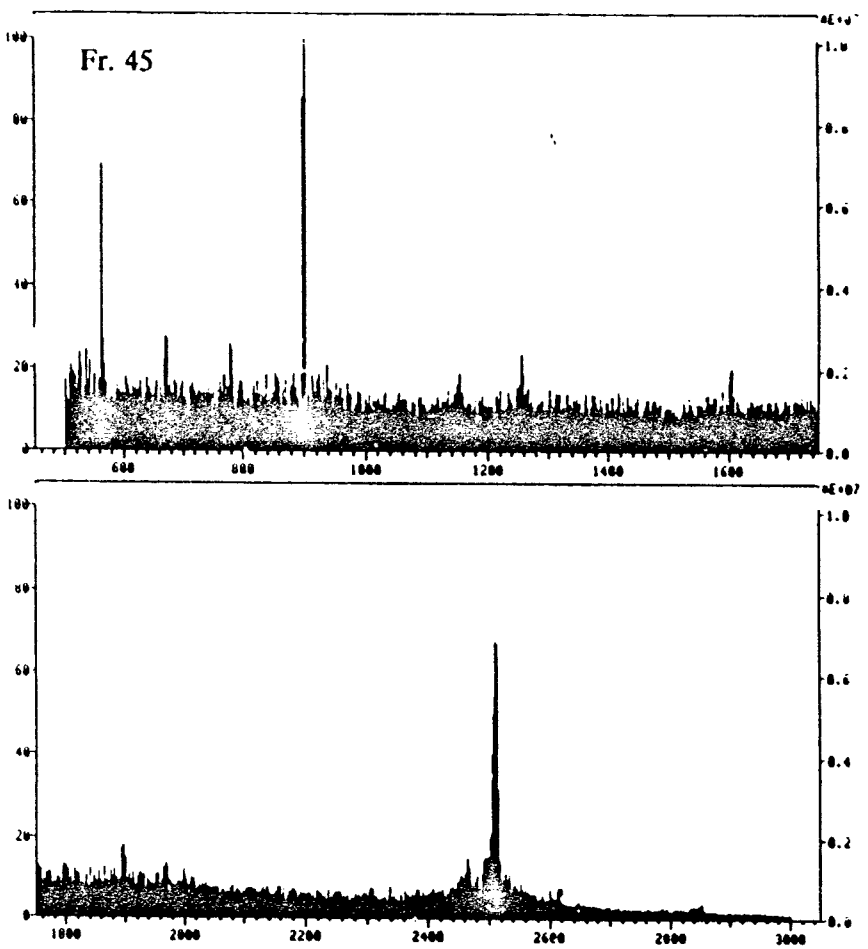

Fr. 40
1889

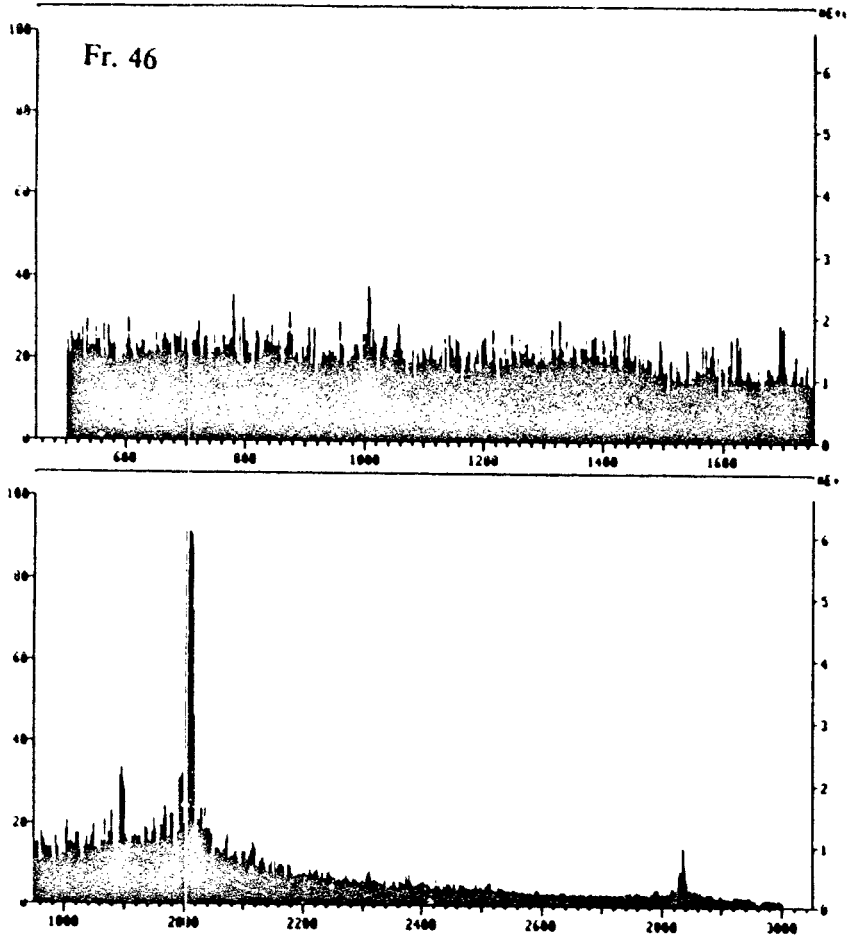

Fr. 41

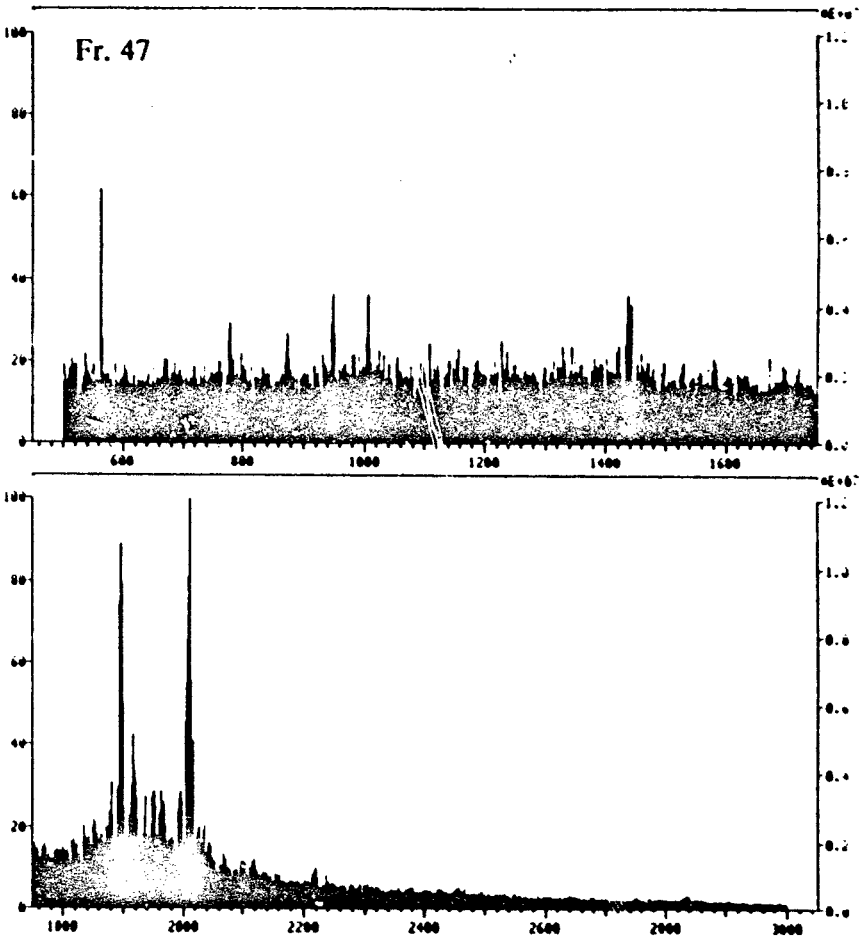

Fr. 42
x4


Fr. 43


Fr. 44


Fr. 45


Fr. 46


Fr. 47

Appendix B

List of additional preliminary sequence information, obtained from various digests. X = Ile or Leu; Z = Gln or Lys; Tr = trypsin; Ch = chymotrypsin; An = protease Asp-N.

m/z	protease	sequence	m/z	protease	sequence
407	Tr	FXK	471	Tr	SXPR
545	Tr	VDAXVK	630	Tr	ZXNZK
547	Tr	TSNXX	622	Tr	YXGXR
644	Tr	GXXTXK	659	Tr	XNVSVK
681	Tr	NYGSXK	690	Tr	TXXESK
701	Tr	NZ(NX)GR	739	Tr	SFNXMK
740	Tr	SFNXMK	750	Tr	FDNXXK
743	Tr	(XN)XEVK	763	Tr	SMXANAR
780	Tr	(--)YXVK	817	Tr	WEEXXK
849	Tr	FXZXVTK	884	Tr	XVGZPTNR
887	Tr	STXXXANR	897	Tr	XXQPXTGR
907	Tr	XYSGXQVK	935	Tr	(QX)SEVMTK
986	Tr	(AR)VSVANXR	1041	Tr	NXWXXPER
1134	Tr	XKSSSVXNMR	1160	Tr	(QA)VVTESXDR
1185	Tr	VVDSDXSXXPK	1196	Tr	DXDTXYETAR
1202	Tr	YGXPVXADXNK	1271	Tr	(DZ)XXXNHGFSK
1416	Tr	(--TTX)SMVPZKR	1526	Tr	ZNZVYXYVVASK
779	Tr	XNFZEK	999	Tr	XXXSYFN/DK
688	Ch	SNXZNX	768	Ch	F/MRHYM
801	Ch	(DZ)AXEXX	812	Ch	NHEXNW
829	Ch	XNEVZNX	896	Ch	XXZPXTGR
1102	Ch	DXZZXENEX	1282	Ch	(PE)XVNZPVZAAX
962	Ch	XZNVTZXF			
732	Tr/An	YGXPVXA	819	Tr/An	DPXFXSK
945	Tr/An	DTGVXSXXK	961	Tr/An	HTHSFVYA
985	Tr/An	DNNTAXXPK	1023	Tr/An	DNVNXVPNK
1471	Tr/An	DXZZXEXEXNZK			